

# UNITED STATES AIR FORCE RESEARCH LABORATORY

---

## TRAINING RESEARCH IN AUTOMATED INSTRUCTION (TRAIN)

J. Wesley Regian  
Barry M. Goettl  
Alan Ashworth III

Air Force Research Laboratory  
Information Systems Training Branch  
2504 Gillingham Drive, Bldg 170  
Brooks City Base, TX 78235-5108

Daniel deBoom  
Michael Anthony  
Katarina Derek

Galaxy Scientific Corporation  
3120 Fire Road  
Egg Harbor Township NJ 08234

Shirley Snooks

Air Force Occupational Measurement Squadron  
1550 5<sup>th</sup> Street  
Randolph AFB TX 78150

**November 2003**

Approved for public release; distribution is unlimited.

AIR FORCE MATERIEL COMMAND  
AIR FORCE RESEARCH LABORATORY  
Human Effectiveness Directorate  
Warfighter Training Research Division  
6030 South Kent Street  
Mesa AZ 85212-6061

## **NOTICES**

**The views expressed in this report are those of the authors and do not necessarily reflect official views of the US Air Force or the Department Of Defense.**

**Using Government drawings, specifications, or other data included in this document for any purpose other than Government-related procurement does not in any way obligate the US Government. The fact that the Government formulated or supplied the drawings, specifications, or other data, does not license the holder or any other person or corporation, or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.**

**The Office of Public Affairs has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.**

**This report has been reviewed and is approved for publication.**

**ALAN ASHWORTH III  
Project Scientist**

**HERBERT H. BELL  
Technical Advisor**

**CURTIS J. PAPKE, Colonel, USAF  
Chief, Warfighter Training Research Division**

**Contract Number: F41624-95-C-5007  
Contractor: Galaxy Scientific Corp.**

Direct requests for copies of this report to:

Defense Technical Information Center  
8725 John J. Kingman Road, Suite 0944  
Ft. Belvoir, VA 22060-6218  
<http://stinet.dtic.mil>

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. <b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</b>					
1. REPORT DATE (DD-MM-YYYY) November 2003		2. REPORT TYPE Final		3. DATES COVERED (From - To) May 1995 to September 2003	
4. TITLE AND SUBTITLE  Training Research in Automated Instruction (TRAIN)				5a. CONTRACT NUMBER F41624-95-C-5007	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER 62205F	
6. AUTHOR(S) J. Wesley Regian; Barry M. Goettl; Alan Ashworth; Dan deBoom; Michael Anthony; Katarina Derek; Shirley Snooks				5d. PROJECT NUMBER 1123	
				5e. TASK NUMBER A3	
				5f. WORK UNIT NUMBER 14	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)  Air Force Research Laboratory Information Systems Training Branch 2504 Gillingham Drive, Bldg 170 Brooks City Base, TX 78235-5318				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory Human Effectiveness Directorate Warfighter Training Research Division 6030 South Kent Street Mesa AZ 85212-6061				10. SPONSOR/MONITOR'S ACRONYM(S) AFRL; AFRL/HEA	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-HE-AZ-TR-2003-0083	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES Air Force Research Laboratory Technical Monitor: AFRL/HEAI, Dr. Alan Ashworth, Comm (210) 536-5535, DSN 240-5535					
14. ABSTRACT This report provides brief research summaries of work performed from May 1995 to September 2003 in support of the Air Force Research Laboratory's Cooperative Laboratory (Co/Lab) Training Research for Automated Instruction (TRAIN) Laboratory.					
15. SUBJECT TERMS Adaptive Instruction; Aptitude-Treatment Interaction; Automated Instruction; Curriculum; Instructional Approach; Intelligent Tutoring System; Mastery Learning; Skill Acquisition; Training Research in Automated Instruction; TRAIN					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT  UNLIMITED	18. NUMBER OF PAGES  68	19a. NAME OF RESPONSIBLE PERSON Ms Liz Casey
a. REPORT UNCLASSIFIED	b. ABSTRACT UNCLASSIFIED	c. THIS PAGE UNCLASSIFIED			19b. TELEPHONE NUMBER (include area code) 480.988.6561 x-188 DSN 474-6188



## Table of Contents

<i>Space Fortress: The Effects of Competition, Skill, and Anxiety on Task Performance</i>	1
<i>The Console Maze: Learning with 2-D vs. 3-D Tasks</i>	2
<i>Space Fortress: Does Observational Learning Affect Performance?</i>	2
<i>Phoenix: Effects of Interference on Acquisition of Automaticity</i>	3
<i>Cooperative Algebra Word Problem-Solving Pilot</i>	4
<i>Space Fortress: The Effects of Discussion, Competition, and Acquisition on Performance</i>	4
<i>Space Fortress: Do Training Differences Influence Resistance</i>	5
<i>to Secondary Task Interference?</i>	5
<i>Space Fortress: Comparison of Focus on Component Scores vs Focus on Total Score on Performance?</i>	6
<i>Variability of Training and its Influence on Performance after Four Hours</i>	6
<i>The Influence of Graphic Representation and Written Cues on Learning a Procedural Task</i>	7
<i>Space Fortress: Verification of Old vs. New Task</i>	7
<i>Phoenix: The Influence of Full Task Training vs. Component Training I</i>	8
<i>Space Fortress: Comparison of Discussion vs. Individual Training on Performance</i>	9
<i>Space Fortress -- Dyadic Training without Social Contact</i>	9
<i>Investigating Backward Transfer and Whole vs. Component Task Training on Knowledge Acquisition</i>	10
<i>Efficiency of Learning: Dyads vs. Individuals for Algebra Word Problems</i>	11
<i>The Influence of Abbreviated vs. Extended Practice in a Two-year Longitudinal Study</i>	11
<i>Stat Lady I -- Probability Module Workbook vs. Computer-Based Tutor</i>	12
<i>Spatial Learning in a Physical vs. Virtual Environment</i>	13
<i>Investigating How Different Learning Environments Affect Exploratory Behavior</i>	13
<i>Intelligent Instruction vs. Passive and Rudimentary Coaching in Knowledge Acquisition</i>	14
<i>The Use of Synthetic Speech in Investigating the Perceptual Effects Within Training</i>	15
<i>Phoenix: Investigating Effects of Blocked vs. Varied Practice</i>	16
<i>The Effectiveness of Computer-Based Word Problem Solving Environment (WPSE) Compared to Paper &amp; Pencil Tests</i>	17
<i>Feedback: The Effect of Elaborative vs. Corrective and Delayed vs. Immediate</i>	18
<i>Aircraft Recognition: The Effect of Random Distributed Practice for Long-Term Training</i>	19
<i>Space Fortress: Longitudinal Investigation of Knowledge Decay and Reacquisition of Skill</i>	20
<i>Phoenix: Gender Differences in Component Task Training</i>	20
<i>Space Fortress: The Effects of Training with Variable vs. Multiple Emphasis on Components</i>	21
<i>Algebra Word Problems (AWP) Dyads vs. Individuals Passive and Interactive</i>	22
<i>Phoenix: Observational vs. Hands-on Learning</i>	23
<i>ACTIVE vs. MITT: Preference of Non-Gaming vs. Gaming Instruction</i>	24
<i>The Effect of Various Types of Explanation on Learning and Retention</i>	24
<i>Testing the Generalizability of the Effects of Massed vs. Distributed Practice</i>	25
<i>The Influence of Physical Exercise on the Effects of Massed vs. Distributed Practice</i>	26

<i>Reducing the Time of Training to Test the Effects of Massed vs. Distributed Practice</i>	27
<i>Phoenix: The Interaction Between Instructional Modality and Gender (Pilot Study)</i>	28
<i>Phoenix: The Interaction between Instructional Modality</i>	28
<i>(Text vs. Text and Speech) and Gender</i>	28
<i>Phoenix: Testing the Effectiveness of Skill Acquisition (Joystick vs. Keyboard)</i>	29
<i>Drive Simulator (Pilot)</i>	30
<i>Stat Lady II: Piloting the Descriptive Statistics Module of the Tutor</i>	30
<i>System T<sup>2</sup>WO: Tutor vs. Traditional Classroom Instruction for Individuals vs. Dyads</i>	31
<i>Team Composition and Type of Worked Examples in Algebra Word Problem Solving</i>	32
<i>Phoenix: Investigating the Optimal Ratio of Practice to Observation for Learning</i>	32
<i>Investigating Three Theories of Symptom Elaboration</i>	34
<i>Comparing Two Software Versions (Pilot Study)</i>	35
<i>Stat Lady II: Piloting the Data Organization and Plotting Module (DS-1) Criterion Tasks</i>	35
<i>Stat Lady II: Testing the Validity of Stat Lady as a Computer-Based Tutor</i>	36
<i>Investigating how Students Learn Troubleshooting Skills</i>	37
<i>ATM 2 -- Task differences in Alternating Task Modules</i>	38
<i>The Effects of Exercise on the Spacing Effect (A Control Group)</i>	38
<i>Stat Lady II -- Descriptive Statistics Module (Computer-Based Tutor) Study D: Predictive Validity of SMART</i>	39
<i>Stat Lady II -- Descriptive Statistics Module (Computer-Based Tutor) Pilot C: Microadaptive Functioning of SMART (Debugging and Preliminary Results)</i>	40
<i>Assessing the Reliability of the Algebra Word Problem Solving (AWP) Pre/Post Tests</i>	41
<i>Testing the Number of Trials to Achieve Automaticity Utilizing the Phoenix and Alphabet Arithmetic Tasks</i>	42
<i>Space Fortress: Assessing Gender Differences with Spatial Tasks</i>	43
<i>Stat Lady II -- Descriptive Statistics Module (Computer-Based Tutor) Study C: Microadaptive Functioning of SMART</i>	43
<i>Early vs. Late Spacing of Training and its Effect on Automaticity</i>	44
<i>Space Fortress: The Effects of Gender, Testosterone Levels, Group Discussion, and Stress on Performance</i>	45
<i>Stat Lady II -- Descriptive Statistics Module (Computer-Based Tutor) Stat Lady E: Replication Study: Differential Practice Effects on Retention</i>	46
<i>Phoenix: Gender Differences Due to Spacing or Observational Effects?</i>	47
<i>Assessing Individual Differences Using Individual Goals</i>	48
<i>Space Fortress: Examining Top-Down Processing</i>	49
<i>Validating Alternative Pedagogical Approaches to</i>	57
<i>Teaching Basic Algebra Skills</i>	57
<i>Building a Bug Library</i>	59

# TRAINING RESEARCH IN AUTOMATED INSTRUCTION (TRAIN)

## Space Fortress: The Effects of Competition, Skill, and Anxiety on Task Performance

**Scientists:** Regian and Shebilske

**Instructors:** Connolly-Gomez and Robertson-Schüle

**Total Subjects:** N = 153

**Purpose of Study:** The experiment was designed to examine the effects of competition on computer-based instruction for trainees with high to low interaction skill and anxiety. Subjects were taught to play Space Fortress and randomly assigned to groups of (1) no competition (lottery determined winner of \$100.00), (2) competition given early, or (3) competition given late in both (early or late competition groups, highest score won \$100.00). In addition, individual difference measures were taken of skill level (high vs. low) and interaction anxiety (high vs. low). There are implications for improving performance in a cost-efficient manner by providing competition between trainees.

**Results of Study:** Overall, subjects in Dyad and Tetrad training groups learned as quickly as subjects in individual training groups. In line with Social Learning Theory, arousal facilitated dominant responses. Thus, early in training, dominant responses are incorrect, but later in training dominant responses are correct. Competition enhanced performance for trainees with high skill level and appears to have no effect on performance with low skill level, although the screening procedure eliminated those trainees with the lowest skill levels. Subjects high in interaction anxiety performed better in Individual protocols, whereas subjects who have low interaction anxiety performed better in Dyad protocols.

### Selected Publications:

Arthur, W., Jr., Young, B., Jordan, J.A., & Shebilske, W.L. (1996). Effectiveness of dyadic training protocols: The influence of training interaction anxiety. *Journal of the Human Factors Society*, 38, 79-86.

Shebilske, W. L., Regian, J. W., & Jordan, J. (1992). A dyadic protocol for training complex skills. *Human Factors*, 34, 369-374.

Worchel, S., Shebilske, W.L., Jordan, J.A., & Prislín, R. (1997). Competition and performance on a computer-based complex perceptual-motor task. *Human Factors*, 39, 410

**Keywords:** Space Fortress, competition, individual, dyad, tetrad, interaction anxiety

## **The Console Maze: Learning with 2-D vs. 3-D Tasks**

**Scientists:** Regian, Shebilske, and Monk

**Instructors:** Connolly-Gomez, Robertson-Schüle, and Monk

**Total Subjects:** N = 31

**Purpose of Study:** To examine how subjects learn a 2-dimensional console maze in a computer-based program and to compare the results to those obtained using a 3-dimensional layout in virtual reality. Issues such as simulation fidelity and God's eye-view perspective were examined.

**Results of Study:** Subjects learned spatial-procedural and spatial-navigational skills. They continued to acquire skill while being tested on the procedure, as the tests provided continued practice in executing the procedure. The maze data showed that the subjects were able to learn the virtual maze, as evidenced by their ability to navigate within the maze after training.

### **Selected Publications:**

Regian, J.W., Shebilske, W.L., & Monk, J. M. (1992). Virtual reality: An instructional medium for visual spatial tasks. *Journal of Communication, Autumn, 42*(4).

Regian, J.W., Shebilske, W.L., & Monk, J.M. (1992). A preliminary empirical evaluation of virtual reality as an instructional medium for visual-spatial tasks. *Journal of Communication, 42*(4), 136-149.

**Keywords:** Space Fortress, 2-dimensional layout, maze, spatial, fidelity

## **Space Fortress: Does Observational Learning Affect Performance?**

**Scientists:** Regian and Shebilske

**Instructors:** Connolly-Gomez and Robertson-Schüle

**Total Subjects:** N = 78

**Purpose of Study:** To determine the effect of passive observation on Space Fortress performance. Observers in the No Observation Tetrad did not observe players during game, but did filler tasks. This experiment has implications for isolating observation factors as opposed to distribution of practice. Tetrad and No Observation Tetrad have same distribution, but different levels of observation.



**Results of Study:** Individual, Dyad, and Tetrad protocols replicated results from previous studies. The No Observation Tetrad performed the same protocol as the Tetrad subjects in Space Fortress. Therefore, distribution of practice rather than observation facilitates performance in Tetrad protocol. However, there was not sufficient experimental control to completely restrict observation. Replication of the results is needed before submission for publication

**Selected Publications:** Awaiting follow-on study.

**Keywords:** Space Fortress, no observation, spatial, cooperative learning, individual, dyad, tetrad

### **Phoenix: Effects of Interference on Acquisition of Automaticity**

**Scientists:** Goettl and Regian

**Instructors:** Connolly-Gomez and Robertson-Schüle

**Total Subjects:** N = 104

**Purpose of Study:** Subjects performed various single and dual tasks while using the Phoenix desktop-flight simulator to determine quantitative and qualitative aspects of interference on cognitive demands. Tasks included a tracking task (highway-in-the-sky) and a limerick task. Performance measures and subjective workload measures were hypothesized to indicate both the overall interference between tasks, and the specific nature of the interference. The goal is to develop a battery of primary and secondary tasks to be used for future studies on the acquisition of automaticity. Length of the study was approximately 6 to 8 hours.

**Results of Study:** This experiment served as a pilot study for the acquisition of the Phoenix task. Limerick interfered minimally with Phoenix.

**Selected Publications:** None.

**Keywords:** Spatial, secondary task, automaticity

### **Cooperative Algebra Word Problem-Solving Pilot**

**Scientists:** Yadrick and Regian

**Instructors:** Connolly-Gomez and Robertson-Schüle

**Total Subjects:** N = 20

**Purpose of Study:** To pilot different problem sets with varying levels of difficulty, instructions, and procedures. Subjects were tested every week for a maximum of one day and a 2-hour slot minimum. This was a pilot for a later study.

**Results of Study:** The study assisted in normalizing problems and established difficulty level.

**Selected Publications:** None.

**Keywords:** Algebra word problems, quantitative

### **Space Fortress: The Effects of Discussion, Competition, and Acquisition on Performance**

**Scientists:** Regian and Shebilske

**Instructors:** Connolly-Gomez and Robertson-Schüle

**Total Subjects:** N = 159

**Purpose of Study:** An earlier Space Fortress competition experiment was replicated with videotaped group discussions added. The emphasis of the study was to determine the effects of group discussion on Space Fortress performance with subjects in Individual training protocol. There are implications for peer-tutoring and cooperative learning contexts during training. This study is linked with a later discussion study.

**Results of Study:** Subjects that were trained under the Individual in Discussion protocol performed significantly better on the Space Fortress task than subjects trained under the Individual protocol. The results suggest that a relatively inexpensive manipulation, such as group discussion can improve performance on complex skill acquisition.

**Selected Publications:**

Prislin, R., Jordan, J.A., Worchel, S., Semmer, F.T., & Shebilske, W.L., (1996). The effects of group discussion on acquisition of complex skills. *Human Factors*, 38, 404

**Keywords:** Space Fortress competition and individual, dyad, tetrad in discussion

**Space Fortress: Do Training Differences Influence Resistance  
to Secondary Task Interference?**

**Scientists:** Regian and Shebilske

**Instructors:** Connolly Gomez and Robertson-Schüle

**Total Subjects:** N = 78

**Purpose of Study:** This experiment was a partial replicate for the Space Fortress competition experiment (pg. 1) (no completion condition). Transfer and secondary tasks were added after lunch on the second day in place of the 2-dimensional console/maze. The secondary tasks were: keyboard, limerick, tapping, protect the fortress, and a hybrid version of the Gopher version of Space Fortress. The purpose was to determine if subjects who learn Space Fortress under different training protocols are more resistant or less resistant to secondary task interference and how effectively they are able to transfer skills compared to their counterparts.

**Results of Study:** We found that the advantages of using dyad and tetrad protocols were maintained throughout the secondary and transfer tasks.

**Selected Publications:**

Shebilske, W.L., Jordan, J.A., Arthur, W., & Regian, J.W. (1993). Combining a multiple emphasis on components protocols for training complex skills. *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting* (1212-1220). Santa Monica, CA: The Human Factors and Ergonomics Society.

**Keywords:** Space Fortress Transfer, secondary tasks, spatial, cooperative learning, individual, dyad, tetrad

## **Space Fortress: Comparison of Focus on Component Scores vs Focus on Total Score on Performance?**

**Scientists:** Shebilske and Regian

**Instructors:** Connolly Gomez and Robertson-Schüle

**Total Subjects:** N = 83

**Purpose of Study:** The goal was to determine the effects that Gopher's Multiple Emphasis on Components (MEC) paradigm would have on subjects who learned Space Fortress under different training protocols. Subjects were instructed to either focus on components of total score (e.g., control of ship, speed of ship, mines, etc.) to raise their score, or on total score (a combination of all those elements) to raise their scores. This has implications for combining Active Interlocked Modeling (AIM) with MEC to create new, more effective training protocols.

**Results of Study:** Gopher's results were replicated as there was an increase in performance with subjects trained under the Individual protocol which focused on the Multi Emphasis Components (MEC). It was also found that MEC combined with Dyad protocol, but not Tetrad protocol, produced more efficient training.

### **Selected Publications:**

Shebilske, W.L., Jordan, J.A., Arthur, W., & Regian, J.W. (1993). Combining a multiple emphasis on components protocols for training complex skills. *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting* (1212-1220). Santa Monica, CA: The Human Factors and Ergonomics Society.

**Keywords:** Space Fortress, multi-emphasis, individual, dyad, tetrad

## **Variability of Training and its Influence on Performance after Four Hours**

**Scientists:** Goettl and Regian

**Instructors:** Connolly Gomez and Robertson-Schüle

**Total Subjects:** N = 70

**Purpose of Study:** This study examined the effects of variability of precision during acquisition in a four-group, between-subjects experimental design. The variable manipulated for this study included the size of the target gates (large or small) during acquisition, transfer, and re-acquisition. It was hypothesized that

variability during acquisition may be more important in transfer. The length of the study was approximately 4-6 hours.

**Results of Study:** Results indicated that including a single, small gate early in the practice phase had an immediate negative effect on performance, but a transient positive effect on later trials.

**Selected Publications:** None.

**Keywords:** Spatial, acquisition, variability, transfer

### **The Influence of Graphic Representation and Written Cues on Learning a Procedural Task**

**Scientists:** Farquhar and Regian

**Instructors:** Connolly Gomez and Robertson-Schüle

**Total Subjects:** N = 153

**Purpose of Study:** To determine whether or not graphic representation, written cues, or both assisted in learning a procedural task.

**Results of Study:** It was found that a graphical model aided in the acquisition of the procedural task.

**Selected Publications:**

Farquhar, J.D., & Regian, J.W. (1993). The effects of a dynamic graphical model during simulation-based training of console operation skill. *Proceedings of the 1993 Conference on Intelligent Computer-Aided Training and Virtual Environment Technology.*

Farquhar, J.D. (1992). A preliminary investigation of the acquisition of console-operation skill through a dynamic graphical model (Technical report submitted to the *Air Force Office of Scientific Research Summer Research Program*).

**Keywords:** Procedural tasks, graphics, learning aids

### **Space Fortress: Verification of Old vs. New Task**

**Scientists:** Shebilske and Regian

**Instructors:** Connolly Gomez and Robertson-Schüle

**Total Subjects:** N = 17

**Purpose of Study:** To determine if Space Fortress performance of subjects trained under the original Gopher version differed from those subjects trained under the new (Galaxy) version of Space Fortress. This will determine if results from past studies can be compared to those of future studies.

**Results of Study:** Results suggest that there are no significant differences between the (Galaxy) version and the Gopher version of Space Fortress. Therefore, previous and future studies using these different versions may be compared without concern.

**Selected Publications:** None.

**Keywords:** Space Fortress, generalizability, individuals

### **Phoenix: The Influence of Full Task Training vs. Component Training I**

**Scientists:** Goettl and Regian

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 83

**Purpose of Study:** This study represented the first attempt of examining whole-/component-task training issues for the Phoenix desktop-flight simulator. Subjects were randomly assigned to one of two groups (full-task training or component task training), acquisition was examined between groups that had received training either presented as a whole task or decomposed as elements of the task. Task elements were presented in a blocked schedule during the training phase. All subjects received a retention test using the whole task rather than the decomposed task. A revision of this pilot experiment was conducted later using three treatment groups with two levels of the component training variable. The length of the study was 16 hours.

**Results of Study:** A factor analysis revealed several component skills important to Phoenix. One factor, "gate aiming" was important early in training. Another factor, "Spatial Orientation" was important early and late in training. This study suggests the utility of backward transfer technique.

**Selected Publications:**

Goettl, B.P. (1993). Analysis of a skill on a flight simulator: Implications for training. *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting* (1257-1261). Santa Monica, CA: The Human Factors and Ergonomics Society.

**Keywords:** Spatial, whole/part training, transfer, acquisition, components, factor analysis

**Space Fortress: Comparison of Discussion vs. Individual Training on Performance**

**Scientists:** Shebilske and Regian

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 103

**Purpose of Study:** To determine the effects of group discussion on Space Fortress performance of subjects in Individual training protocol. This has implications for peer-tutoring and cooperative learning contexts during training. This was an extension of the previous Discussion study.

**Results of Study:** Subjects that were trained under the Individual in Discussion protocol performed significantly better on the Space Fortress task than subjects trained under the Individual protocol. The results suggest that a relatively inexpensive manipulation such as group discussion can improve performance on complex skill acquisition.

**Selected Publications:**

Prislin, R., Jordan, J.A., Worchel, S., Semmer, F.T., & Shebilske, W.L., (1996). The effects of group discussion on acquisition of complex skills. *Human Factors*, 38, 404  
Prislin, R., Jordan, J. A., Shebilske, W.L., (In preparation).

**Keywords:** Space Fortress, individual, individual in discussion

**Space Fortress -- Dyadic Training without Social Contact**

**Scientists:** Shebilske and Regian

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 86

**Purpose of Study:** To determine if the partner in the Dyad training protocol could be replaced with a "computer" partner. Also, to isolate and control components (e.g., social motivation) which contribute to the effectiveness of dyadic training. Subjects were told that the computer was controlling the other half of the game when in reality it was their anonymous partner--another subject who was playing the other half of the game. This has implications for whether social contact is necessary as long as the subject is able to model the actions and reactions of some type of partner (even if it is a computer).

**Results of Study:** Subjects that were trained under the Hidden-Partner Dyad performed as well as subjects trained under Individual protocol and Dyad protocol from previous studies. This suggests that social contact is not necessary as long as modeling of response can occur. However, training two subjects at one terminal still has a greater cost-benefit ratio.

#### **Selected Publications:**

Shebilske, W.L., Regian, J.W., Arthur, W., & Jordan, J. (1992). A dyadic protocol for training complex skills. *Human Factors*, 34, 369-374.

**Keywords:** Space Fortress, Dyads without social interaction, spatial

### **Investigating Backward Transfer and Whole vs. Component Task Training on Knowledge Acquisition**

**Scientists:** Goettl and Regian

**Instructors:** Connolly Gomez and Robertson-Schüle

**Total Subjects:** N = 60

**Purpose of Study:** This follow-up study examined task component vs. whole task training using the Phoenix desktop-flight simulator. Subjects were assigned to one of three groups: (1) whole task training, (2) component skills training with whole task transfer, or (3) component skills training control. Acquisition was examined between groups that had received training either on task components that were presented as a whole task or decomposed as elements of the task. Both transfer and knowledge tasks were used to assess acquisition during training. The approximate length of the study was 16 hours.

**Results of Study:** Replication of Phoenix: The Influence of Full Task Training vs. Component Training; the factor analysis identified two component skills important to Phoenix. One factor, "gate aiming," was important early in training. Another factor, "Spatial Orientation," was important early and late in training. This



study verifies the results found in the previous study and suggests the utility of backward transfer technique.

**Selected Publications:** None.

**Keywords:** Spatial, whole/part training, transfer, components, acquisition, factor analysis

### **Efficiency of Learning: Dyads vs. Individuals for Algebra Word Problems**

**Scientists:** Yadrick and Regian

**Instructors:** Connolly Gomez and Robertson-Schüle

**Total Subjects:** N = 30

**Purpose of Study:** To compare teaming differences using the unguided version of Mtutor. The subjects were placed in either an individual or dyad condition using the unguided version of the Mtutor.

**Results of Study:** There were no differences found for dyads or individuals. There was no learning effect, due to ceiling effects and minimal learning had taken place.

**Selected Publications:** None.

**Keywords:** Algebra word problems, quantitative, cooperative learning, dyads vs. individuals

### **The Influence of Abbreviated vs. Extended Practice in a Two-year Longitudinal Study**

**Scientists:** Shute and Gawlick-Grendell

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 92

**Purpose of Study:** To compare retention of information acquired from the Flight Engineering Tutor training acquired two years previously. In the original study (1990), there were different practice conditions where students would receive either three abbreviated or 12 extended practice problems. In addition, sensitivity to relearning was investigated following a 2-hour refresher course.

**Results of Study:** Two experiments examined practice effects on parameters of skill acquisition, learning outcome, retention, and sensitivity to re-learning. In Experiment 1, our learning criterion task was an intelligent tutoring system teaching flight engineering knowledge and skills with two practice conditions: "Extended"

and “Abbreviated,” differing only in the number of practice problems requiring solution across problem sets. Results from the first experiment showed that subjects in the Abbreviated condition completed the curriculum, overall, in half the time it took the Extended subjects, but at the expense of greater errors and latencies during problem solution. Despite these acquisition differences, both groups performed the same across *all* learning outcome measures. Experiment 2 examined the benefits of extended practice on retention and sensitivity to relearning after a significant period of time had elapsed (2 years). Although the sample size was fairly small (34 returning subjects), we found significant group differences in terms of retention of previously learned material. Following a 2-hour “refresher course,” there was no evidence of differential sensitivity to re-learning. Data showed that the benefits of initial extended practice emerged.

### **Selected Publications:**

Shute, V.J. & Gawlick-Grendell, L.A. (1994, April). Practice effects on skill acquisition, outcome, and retention. Paper presented at the *American Educational Research Association*, New Orleans, LA.

Shute, V.J. & Gawlick L.A. (1995). Practice effects on skill acquisition, outcome, and retention. *Human Factors*, 12, 1.

**Keywords:** Flight Engineering Tutor, retention

### **Stat Lady I -- Probability Module Workbook vs. Computer-Based Tutor**

**Scientists:** Shute and Gawlick-Grendell

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 107 computer-based, N = 104 workbook version, N = 100 no treatment control

**Purpose of Study:** The purpose of the study was twofold: (1) to test the computer's contribution to learning outcome, efficiency, and overall enjoyment of the learning process and  
(2) to compare the possible differences in instruction via statistics.

**Results of Study:** Overall, there were no significant differences between treatment conditions (e.g., Stat Lady computer versus workbook). However, students placed in the computer-based condition learned more declarative knowledge in comparison to students placed in the workbook condition. Aptitude-treatment interactions were also investigated, and the results found that high-aptitude students

learned more from the computer version of the tutor, while low-aptitude students learned the same, regardless of treatment condition.

**Selected Publications:**

Shute, V.J., Gawlick-Grendell, L. A., Young, R.K., & Burnham, C.A. (1995). An experiential system for learning probability: *Stat Lady* description and evaluation. *Instructional Science* (4) 1-22.

**Keywords:** Stat Lady - Intro. to Probability Statistics Tutor, ITS, ATI

**Spatial Learning in a Physical vs. Virtual Environment**

**Scientists:** Regian, Shebilske, and Monk

**Instructors:** Monk, Morlen, and Gonzalez

**Total Subjects:** N = 22

**Purpose of Study:** The purpose was twofold: (1) to investigate differences in spatial learning between a VR environment and the actual physical environment, and (2) to describe several applications of Virtual Reality technology in conjunction with the tasks for the Federal Bureau of Investigation (FBI) during the 1996 Olympic Games.

**Results of Study:** The results were that the subjects were able to navigate through the physical environment after training on the 2-D console and maze.

**Selected Publications:**

Regian, J.W. (1986). *An assessment procedure for configurational knowledge of large-scale space*. Unpublished doctoral dissertation. Santa Barbara CA: University of California.

Regian, J.W. & Shute, V.J. (1993, April). An investigation of virtual reality for training visual-spatial tasks. Paper presented at the *American Educational Association*, Atlanta, GA.

**Keywords:** Virtual reality versus Physical, spatial

**Investigating How Different Learning Environments Affect Exploratory Behavior**

**Scientists:** Shute

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 400

**Purpose of Study:** The main question addressed whether there is a main effect of learning environment when comparing learning outcomes between discovery (or inductive) and directed kinds of instructional approaches, furthermore, can aptitude-treatment interactions (ATI) be replicated where some people learn better in discovery environments and others learn better under different conditions? The study investigated learning differences as a function of being assigned to either a rule-induction or rule-application learning environment using a confirmatory test of ATI. That is, subjects were placed *a priori* in one of two computerized learning environments based on a decision rule obtained from a previous study using an electricity tutor (Shute, 1993). The learning style (or aptitude) investigated was "exploratory behavior."

**Results of Study:** A significant 3-way interaction was found among learning environment, exploratory behavior, and math skill. Subjects in the "few" math courses category clearly showed the predicted disordinal interaction between exploratory behavior (low vs. high) and learning environment (inductive vs. applied). However, subjects who had taken "many" math courses showed a disordinal interaction, but in the opposite direction. Implications of these findings are discussed in relation to macroadaptive approaches to the design of ITS.

#### **Selected Publications:**

Shute, V.J. (1992). Aptitude-treatment interactions and cognitive skill diagnosis. In J.W. Regian & V.J. Shute (Eds.), *Cognitive approaches to automated instruction* (pp.1-14). Hillsdale, NJ: Lawrence Erlbaum Associates.

Shute, V.J. (1993). A comparison of learning environments: All that glitters...In S.P. Lajoie & S.J. Derry (Eds.), *Computers as cognitive tools* (pp. 47-74), Hillsdale, N.J: Lawrence Erlbaum Associates.

Shute, V.J. (1994). Learners, environments, and a confirmatory test of ATI. Paper presented at *American Educational Research Association*, New Orleans, LA.

**Keywords:** Electricity Tutor, ITS, induction vs. application, exploratory behavior

#### **Intelligent Instruction vs. Passive and Rudimentary Coaching in Knowledge Acquisition**

**Scientists:** Shahidi and Regian

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 105

**Purpose of Study:** This study investigated the issue of whether students learn better with online intelligent coaching systems driven by student modeling rather than with passive online help systems. There were three treatment groups: (1) passive online help, (2) intelligent instruction based on student model, and (3) rudimentary coaching. It was hypothesized that intelligent instruction using a student model would train subjects better in terms of achievement as measured by the number of steps to solve the problem, total errors, number of critical errors, testing time and time necessary to train. All groups received 3-4 hours of classroom instruction on how to use an electronic spreadsheet and were then given several practice problems and two test problems. The total length of the study was 6-8 hours.

**Results of Study:** Results indicated that the intelligent coaching instruction was significantly better than the rudimentary and passive online coaching systems. Specifically, students trained with the intelligent help using the student model used fewer steps to complete the goal, committed fewer errors, less critical errors, took less time to train and test, and preferred that type of instruction overall.

**Selected Publications:** None.

**Keywords:** ITSs, coaching, instruction, student models, help-systems, quantitative

### **The Use of Synthetic Speech in Investigating the Perceptual Effects Within Training**

**Scientists:** Connolly Gomez, Gonzalez, Shebilske, and Regian

**Instructors:** Connolly Gomez and Robertson-Schüle

**Total Subjects:** N = 133

**Purpose of Study:** Previous studies have revealed that the perception and comprehension of synthetic speech may be attributed to increased processing demands in short-term memory. This increase in short-term memory demand is reflected in serial recall and variable inter-stimulus rate tasks. Additionally, it has been consistently shown that the perception of synthetic speech improves with moderate amounts of training. The present study was conducted to determine if the increased perceptual effects of training are due to a reduction of short-term memory load as reflected in serial recall and varied presentation rate recall tasks. Two groups of subjects were tested with synthetic speech using the same recognition tasks recognition before and after training. One group was trained with synthetic speech and the other group acted as the control, receiving no training between the pretest and posttest interval. The approximate length of the study was 16 hours over a one-week period.

**Results of Study:** Results indicated that subjects who trained on synthetic speech versus the control group, who received no training between baseline and posttest, performed better on short-term memory tasks. The results of this study indicate that training may have reduced some of the cognitive burden that synthetically generated auditory displays place upon the user.

**Selected Publications:**

Connolly Gomez, C.A., Shebilske, W.L., & Regan, J.W. (1994). The effects of training on capacity demands in short-term memory for synthetic speech. *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting* (1229-1233). Santa Monica, CA: The Human Factors and Ergonomics Society.

**Keywords:** Synthetic speech, working memory, intelligibility, capacity demands

**Phoenix: Investigating Effects of Blocked vs. Varied Practice**

**Scientists:** Goettl and Regan

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 82

**Purpose of Study:** This study compared Carlson's working memory hypothesis and Glenberg's multiple-components theory of contextual interference effect. Both theories predict that blocked practice on the task component will be less effective than varied practice. Furthermore, Glenberg predicts that a completely random schedule will be more effective than a schedule consisting of replications of one sequence. Subjects were randomly placed into one of three treatment groups: (1) whole-task training, (2) blocked presented component training, or (3) repeated sequence component training. Subjects were given a retention task on the whole task and a transfer to another task. In addition, spatial and attentional individual differences were assessed using CAM, VSAT, and Space Fortress aiming task. Phoenix testing phases always followed a 24 hr. retention interval. The length of the study was approximately 20 hours.

**Results of Study:** Contextual interference condition showed no benefits over blocked condition. The beneficial effects of mixed practice observed in simple tasks were not obtained in this complex task. These results challenge the generalizability of the effect.

**Selected Publications:**

Goettl, B.P. (1994). Contextual interface effects on acquisition and transfer of a complex motor task. *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting*. (1220-1224). Santa Monica, CA: The Human Factors and Ergonomics Society.

**Keywords:** Spatial, contextual interference, massed practice, whole/part training

**The Effectiveness of Computer-Based Word Problem Solving Environment (WPSE)  
Compared to Paper & Pencil Tests**

**Scientists:** Yadrick and Regian

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 71

**Purpose of Study:** The goal of this study was to compare problem solving skill in a computer-based Word Problem Solving Environment (WPSE) with paper and pencil tests. There were 5 conditions: (1) subjects using WPSE only, (2) subjects using paper and pencil only with worked examples, (3) subjects using "paper WPSE ", (e.g., all problems, instruction screens, hints and help in booklet form), (4) subjects given paper and pencil problems and solutions (they had to work out the problem using the WPSE), and (5) subjects given similar worked example problems, but had to work and enter solutions into the WPSE.

**Results of Study:** Results are complex, but briefly, Groups 1 and 5 did poorly, whereas Groups 2, 3, 4 did relatively well.

**Selected Publications:**

Yadrick, R. M., Regian, J. W., Robertson-Schüle, L., & Gomez, C. C. (1996). Interface, instructional approach, and domain learning with a mathematics problem-solving environment. *Computers in Human Behavior*, 12(4), 527-548.

**Keywords:** Algebra word problems, varied forms of instructional support

## **Feedback: The Effect of Elaborative vs. Corrective and Delayed vs. Immediate**

**Scientists:** Farquhar and Regian

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 128

**Purpose of Study:** Loader is a complex problem-solving task requiring the operation of a remote crane control arm to load various canisters from a set of storage bins to one or more railroad cars via a simulated control-panel console. The computer-based simulation demonstrates a number of features including an on-screen dynamic display, an expert system module, flexible feedback messages, and a student module with the ability to output extensive data for analysis. The experiment was a 2-variable design examining elaborative “intelligent” versus corrective “dumb” feedback and delayed versus immediate feedback. The approximate length of the study was 4-6 hours.

**Results of Study:** Results showed that treatment groups were equivalent in terms of acquisition time however, the elaborative feedback treatment group were 25% faster and 40% more accurate than the corrective feedback group during initial training. These differences disappeared during later training stages. Although subjects indicated their preference for delayed feedback over immediate feedback for non-critical errors, no treatment differences were found suggesting that the tutor may be just as effective when feedback is provided for only a limited number of critical elements.

### **Selected Publications:**

Farquhar, J.D., & Regian, J.W. (1994). The type and timing of feedback within an intelligent console-operations tutor. *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting* (1225-1228). Santa Monica, CA: The Human Factors and Ergonomics Society.

Farquhar, J.D. and Regian, J.W. (1994). Effects of knowledge representation during computer-based training of console operation skill. *Paper presented at the 1994 Conference of the Association for Educational Communications and Technology*, p. 2-15.

Farquhar, J.D. (1993). The design, development, and preliminary evaluation of LOADER: A tool for investigating the acquisition of console-operation skill. *Final Report, Graduate Student Research Program, Armstrong Laboratory*, p. 17-1; 17-15.

Farquhar, J.D. (1993, June). Acquisition of console-operation skill through a dynamic graphical model. *Insight*, 15(2), 5-7.

**Keywords:** Procedural, instructional strategies, feedback, displays, simulation



## **Aircraft Recognition: The Effect of Random Distributed Practice for Long-Term Training**

**Scientists:** Goettl and Regian

**Instructors:** Connolly Gomez and Robertson-Schüle

**Total Subjects:** N = 97

**Purpose of Study:** The contextual interference effect has been well documented in the literature for recognition of complex visual stimuli. However, many of these experiments allowed only a few seconds for "study" time and tested recall minutes after practice. Furthermore, random-distributed practice was shown to produce better recognition than massed practice when testing conditions were varied from practice conditions. The goal of this research is to examine the random-distributed effect for long-term training. Subjects viewed images of military aircraft in one of four conditions: (1) massed, (2) successive-distributed, (3) recursive-distributed, and (4) random-distributed. During each study-test phase, subjects viewed 7 target aircraft for 8 seconds each. Targets were randomly tested and interspersed with 13 distracters. In addition, half of the targets and distracters were shown in a different format (actual pictures). Subjects indicated whether they had seen the aircraft before and rate how confident they were in their answer. The approximate length of the study was 2-3 hours.

**Results of Study:** Benefits of mixed presentation was obtained compared to massed presentations. However, the random presentation (high-contextual interference) was not superior to repeated sequence (low-contextual interference). These results challenge Glenberg's theory of multiple components training.

### **Selected Publications:**

Goettl, B.P. (1996). The spacing effect in aircraft recognition. *Human Factors*, 38, 1.

**Keywords:** Massed practice, distributed practice, visual stimuli, contextual interference

## **Space Fortress: Longitudinal Investigation of Knowledge Decay and Reacquisition of Skill**

**Scientists:** Regian and Shebilske

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 21

**Purpose of Study:** The focus was to determine the effects of a long period of nonpractice (> 6 months) on Space Fortress skill retention and reacquisition. This may have implications for maintenance training in operational training environments and time required to reacquire previously learned skills.

**Results of Study:** As expected, we found substantial forgetting of skills, but also substantial increase in speed of reacquisition of skills. This study serves to establish a baseline for future studies as well as provide important information on maintenance training of complex skills.

**Selected Publications:** None.

**Keywords:** Space Fortress, retention, spatial, individual

## **Phoenix: Gender Differences in Component Task Training**

**Scientists:** Goettl and Regian

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 114

**Purpose of Study:** A factor analysis of Phoenix desktop-flight simulator data (collected in the component skills studies) revealed two major groups of component skills. These two groups of skills are referred to as spatial orientation and gate aiming. Comparison of high-ability and low-ability subjects indicated that high-ability subjects were significantly better on the spatial orientation skills before and after slalom training. However, high-ability subjects performed better than low-ability subjects on gate-aiming skills before, but not after, slalom training. Presumably spatial-orientation skills represent strong individual differences that are difficult to train, but are related to skilled performance on Phoenix. It is hypothesized that gate-aiming skill seems to be a task-specific ability that is easily learned, but not highly predictive of skilled performance on Phoenix. This study compares three training programs: whole task training, spatial-orientation training, and gate-aiming training.

Training effectiveness will be assessed using a part-task paradigm. The length of the study was approximately 16 hours.

**Results of Study:** This study directly tested the importance of these tasks using a part-task training paradigm. One group of participants received part-task training concentrating on the critical component tasks. A second group received part-task training on non-critical component tasks. The third group received whole-task training. Participants receiving part-task training on the critical tasks performed as well as participants in the whole-task condition. Females, but not males, receiving practice on the “non-critical” tasks performed worse than females in the other two groups. This suggests an aptitude by treatment interaction. Results were not replicated on a transfer task. These data illustrate the importance of selecting component tasks for part-task training, and the utility of the backward-transfer technique in identifying those tasks.

#### **Selected Publications:**

Goettl, B.P. (1995). Part-task training of complex tasks: Utility of backward transfer. *Proceedings of the 39th Annual Meeting of the Human Factors & Ergonomics Society* (1345-1349). Santa Monica, CA: The Human Factors & Ergonomics Society.

**Keywords:** Spatial skill, aptitudes, factor analysis, whole/part training

#### **Space Fortress: The Effects of Training with Variable vs. Multiple Emphasis on Components**

**Scientists:** Shebilske, Corrington, Jordan, and Regian

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 118

**Purpose of Study:** Literature on motor skill learning indicates that emphasizing multiple components over trials results in higher acquisition and performance. This finding is somewhat analogous to the part/whole training theory. This suggests that training improves when the task components are practiced individually rather than the entire task practiced as a whole. In this study, a factorial combination of two variables, (Individuals versus Dyads) and (Multiple Emphasis on Components-MEC versus Variable Emphasis on Components-VEC) created 4 between-subject groups. Both MEC and VEC treatment groups receive a total of 16 emphasis for each component score, however the MEC treatment group receives the emphasis on successive game sessions. Conversely, the VEC group receives emphasis instructions that vary within game sessions. In order to enhance the potential immunization against

habituation in massed trials, every adjacent trial within a practice session has a different emphasis. The approximate length of the experiment was 24 hours.

**Results of Study:** It is hypothesized that VEC will share MEC's advantages of part practice within the whole task learning and will have the added advantage of immunizing against degraded performance in massed practice. VEC should also combine effectively with dyadic protocol, which would make VEC-Dyadic protocol the most effective to date for training Space Fortress. Initial results however, do not support this hypothesis.

**Selected Publications:** Follow-up study in preparation.

Shebilske, W.L., Jordan, J., Arthur, W., & Regian, J.W. (1993). Combining a multiple emphasis on components with small group protocols for training complex skills. *Proceedings of the Human Factors and Ergonomics Society, 37th Annual Meeting* (1215-1219). Santa Monica, CA: The Human Factors and Ergonomics Society.

**Keywords:** Variable emphasis, multiple emphasis, whole/part training, dyadic

### **Algebra Word Problems (AWP) Dyads vs. Individuals Passive and Interactive**

**Scientists:** Yadrick and Regian

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 68

**Purpose of Study:** This was a 2 X 2 factorial design where performance differences were compared in solving problems using a passive version of a math tutor (WPSE) versus an active version (Solver). Subjects were assigned to individual or dyad protocol.

**Results of Study:** Individuals showed substantial improvement with the WPSE system, but no improvement with the Solver system. Dyads showed no main effects regarding version; however, they showed moderate improvement from pre to posttest with both versions.

**Selected Publications:**

Yadrick, R. M., Regian, J. W., Robertson-Schüle, L., & Gomez, C. C. (1996). Interface, instructional approach, and domain learning with a mathematics problem-solving environment. *Computers in Human Behavior, 12*(4), 527-548.

Yadrick, R.M., Regian, J.W., Connolly-Gomez, C., & Robertson-Schüle, L. (1997). Dyadic vs. individual practice with exploratory and directive mathematics tutors. *Journal of Educational Computing Research*, 17(2), 165-186

**Keywords:** Algebra Word Problems, dyads vs. individuals

### **Phoenix: Observational vs. Hands-on Learning**

**Scientists:** Goettl, Connolly and Regian

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 88

**Purpose of Study:** Examined which aspects of a complex, psychomotor task necessary for the Phoenix desktop-flight simulator can be learned through observation and what the optimal ratio of observation to hands-on practice is. Subjects were assigned to one of five treatment groups: (1) perform slalom and strike, (2) observe slalom and strike, (3) perform slalom and observe strike, (4) observe slalom and perform strike, and (5) filler task control. The filler task consisted of playing Windows Solitaire. Questionnaires were administered at several points throughout the training phase in order to assess subjects' knowledge of and strategies for Phoenix. The length of the study was approximately 16 hours.

**Results of Study:** This study tested the prediction that observational learning will be more effective for motor tasks having substantial cognitive demands than for those that do not. Subjects were divided into three treatment groups: performers, observers, and no-observe controls. In Phase I, subjects were trained on a computer-based flight task requiring relatively little cognitive demands. In Phase II, subjects were trained on a different flight task that had significant cognitive and strategic demands. In Phase I, performers were superior to both observers and controls; the observers did not differ significantly from the controls. In Phase II, observation showed a beneficial effect for females. The female observers performed as well as the female performers. The results of this study suggest that observational learning benefits tasks with significant cognitive components more than tasks that are primarily psychomotor. Implications for computer-based training are discussed.

### **Selected Publications:**

Goettl, B.P., & Gomez, C.C. (1995). The role of observational learning in automated instruction of complex tasks. *Proceedings of the 39th Annual Meeting of the Human Factors & Ergonomics Society* (1335-1339). Santa Monica, CA: The Human Factors & Ergonomics Society.

**Keywords:** Spatial, observation, cognitive strategies, modeling

### **ACTIVE vs. MITT: Preference of Non-Gaming vs. Gaming Instruction**

**Scientists:** Connolly, Norton, Gonzalez, Kline, Johnson, and Robertson-Schüle

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 73

**Purpose of Study:** The goal of the study was to assess the user preferences and opinions, difference in context-specific knowledge between users, and the value of gaming versus non-gaming computer-based training. Participants were randomly assigned to two treatment groups; MITT (non-gaming) or ACTIVE (gaming). Participants in each treatment group were instructed for approximately 3 hours and then given a posttest and an attitude questionnaire. Following the posttest participants were instructed on the opposite tutor for 1 hour. Participants were then given a second posttest and attitudinal questionnaire. Participants were not used in the study if they rated themselves as experts in automobile maintenance or had never driven a car. The length of the study was approximately 6-8 hours.

**Results of Study:** No statistically significant differences were found for automotive maintenance knowledge between treatment groups following either posttest despite almost 1 hour less time on training for the gaming environment. However, approximately 83% of items were ranked higher by gaming participants on the first attitude questionnaire. After training in both environments a majority of participants preferred and felt they learned more in the gaming environment. In addition, all levels of participants benefited equally and all categories/levels of knowledge were equally acquired from the gaming environment.

### **Selected Publications:**

Norton, J., Gonzalez, G.C., Connolly-Gomez, C., Kline, K.B. & Regian, J.W. (1995). Use of gaming for intelligent technical training (AL-TP-1994-0028, AD B198812). Brooks AFB TX: Armstrong Laboratory.

**Keywords:** Gaming, maintenance, training, automotive, troubleshooting

### **The Effect of Various Types of Explanation on Learning and Retention**

**Scientists:** Hale and Regian

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 73

**Purpose of Study:** This study investigated issues surrounding how participants learn to carry out a simple form of troubleshooting using a fictional industrial system called a glass brick maker (GBM). Participants were presented five diagrams which characterized the system. The diagrams were presented individually and reflect a hierarchy comprised of top-level system diagram followed by four diagrams representing subsystems (which decompose from the top level). Based on the treatment group, subjects received either explanatory information describing the functionality, mechanism, or no explanation of each diagram component. In the learning phase, participants studied and recalled simple diagrams. In the troubleshooting phase, subjects were required to generate fault hypotheses in response to symptoms. Retention tests were given 24, 48, or 96 hours following the learning phase.

**Results of Study:** In this study the component effects model was contrasted with compilation models, which assume that explanations will affect performance prior to rule compilation, but not after compilation has been accomplished. Learners received a rule learning phase immediately following system learning which continued until compilation was successful. Then they were given a re-test 24, 48, or 96 hours later. Explanations affected performance in early trials during both immediate learning and re-testing. This result contradicts predictions of the compilation model, but supports the component effects model. Apparently, explanations are used by learners only for hypothesis generation, prior to accrual of associative strength associated with compilation.

**Selected Publications:** None

**Keywords:** Troubleshooting, diagnosis, feedback, scheduling

### **Testing the Generalizability of the Effects of Massed vs. Distributed Practice**

**Scientists:** Goettl, Yadrick, Connolly, and Regian

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 43

**Purpose of Study:** Distributed practice is overwhelmingly supported in the literature over massed practice for a variety of tasks. However, most of the published studies used simplistic tasks and conducted the practice phase in a matter of minutes or hours. The issue as to whether or not these findings will generalize to real-world training settings, where the tasks are often complex and training is

conducted over several days to weeks is the primary focus of this study. The effects of massing versus distributing practice across 40 hours of training were investigated using three tasks: Space Fortress, Phoenix and Math Word Problems (using paper and pencil). There were two treatment groups that were used and randomly assigned to different weeks; massed practice and distributed practice. To minimize negative transfer between Space Fortress and Phoenix, subjects were instructed only to use the keyboard for Phoenix. In addition, to minimize any massing effects during the distributed practice, tasks were rotated such that no one task was ever practiced successively. Retention and transfer tests always followed 24 hours after practice sessions. The overall length of the study was approximately 40 hours.

**Results of Study:** A significant effect was found for distributing practice Space Fortress and Math Word Problems; performance on Phoenix was not altered. This study was replicated in the new laboratory with Space Fortress program modifications being made.

**Selected Publications:** None.

**Keywords:** Massed/distributed practice, spatial, problem solving

### **The Influence of Physical Exercise on the Effects of Massed vs. Distributed Practice**

**Scientists:** Goettl, Yadrick, Connolly, and Regian

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 85

**Purpose of Study:** The hypotheses were the same as the earlier study; with the addition of a new variable-exercise. Again, the effects of massing versus distributing practice across 40 hours of training were investigated using three tasks: Space Fortress, Phoenix, and Math Word Problems (using paper and pencil). Subjects were placed into one of four conditions: massed, massed with exercise, distributed, and distributed with exercise. It was hypothesized that exercise would offset possible fatigue and boredom effects found in massed practice and improve performance. A pilot study examining different exercises (marching, walking, isometrics, etc.) revealed that a 5-min march increased heart rate on average by 25% and was selected as the exercise to use in the study. The overall length of the study was approximately 40 hours.

**Results of Study:** Results for Space Fortress and Math tasks showed a significant effect for distributed training over massed training. Performance on Phoenix was not altered. Comparisons between the buildings for the same treatment groups also showed a significant effect for Space Fortress with the new program



modifications being made. These results help generalize traditional literature findings of massed versus distributed practice based on relatively simple tasks to more complex tasks which are typical real-world training environments. In addition, this study will help shed light onto the factors responsible for the spaced practice effect.

**Selected Publications:**

Goettl, B.P., Yadrick, R.M., Connolly Gomez, C., Regian, J.W., & Shebilske, W.L. (1996). Alternating task modules in training complex tasks. *Human Factors*, 38, 330-346.

**Keywords:** Massed, spaced, distributed, training, practice, exercise

**Reducing the Time of Training to Test the Effects of Massed vs. Distributed Practice**

**Scientists:** Goettl, Yadrick, Connolly Gomez, and Regian

**Instructors** Connolly and Robertson-Schüle

**Total Subjects:** N = 48

**Purpose of Study:** This experiment was based on previous findings, which indicated strong effects for distributing over massing practice for three different types of tasks: Space Fortress, Phoenix Desktop-Flight Simulator, and Math Word Problems (using paper and pencil format). This study mirrored the previous study in terms of scheduling, however, training time was reduced by approximately 50% and exercise was not manipulated. Subjects in this study received approximately 5 hours of training on each of the tasks: Space Fortress, Math Tutor, and Phoenix. The schedule was arranged so that those subjects in the distributed group would never receive more than 35 minutes of consecutive hours of training on one task. The massed group received the same total amount of training on each task as the distributed group; however, they finished all the training on one task before going on to the next task. Posttests always followed 24 hours after training. Again, to reduce possible negative transfer effects between Space Fortress and Phoenix, subjects were only allowed to use the keyboard, rather than the joystick on the Phoenix task. This study should help us bridge our results in the large massed versus distributed study which contained tasks over a one week period to the literature which used tasks within a very short period of time (usually only a few hours). The approximate length of the study was 20 hours.

**Results of Study:** Currently being analyzed.

**Selected Publications:** *None.*

**Keywords:** Massed, spaced, distributed, training, practice, exercise

### **Phoenix: The Interaction Between Instructional Modality and Gender (Pilot Study)**

**Scientists:** Regian, Goettl, and Connolly  
**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 10

**Purpose of Study:** The effectiveness of instructions presented via text or text and speech and how attitudinal variables may interact with the presentation modality was investigated. It was hypothesized that speech in addition to the text may provide a more motivational environment in which to learn Phoenix, furthermore, it was suggested that there may be an interaction between the instructional modality and gender. Specifically, female subjects may benefit more from the text plus verbal instructions than do male subjects. The length of the experiment was approximately 16 hours spread across 4 days.

**Results of Study:** Not enough subjects (5 per group) completed the study to run any meaningful ANOVAs. However, empirical analysis of the data reveals that the speech group performed worse. This effect may be an artifact of the low number of subjects who finished the study, the unbalanced gender ratio between treatment groups, and differences in the pretest ability. The subjects with speech tended to agree that: (a) the speech made the task more interesting, (b) speech forced them to look at the displays, and (c) they learned more from the speech than if they had just read the screen. None of the subjects agreed with the statement that the speech was hard to understand. In sum, the data we have are not conclusive one way or the other, but if the trend continues with additional data, it would be a very interesting, counter-intuitive finding. A follow-up study is needed to verify these results.

**Selected Publications:** None.

**Keywords:** Modality, training, synthetic speech, spatial, gender

### **Phoenix: The Interaction between Instructional Modality (Text vs. Text and Speech) and Gender**

**Scientists:** Regian, Goettl, and Connolly  
**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 82

**Purpose of Study:** The effectiveness of instructions presented via text or text and speech and how attitudinal variables may interact with the presentation modality was investigated. It was hypothesized that speech in addition to the text may provide a more motivational environment in which to learn Phoenix. It was also hypothesized that there may be an interaction between the instructional modality and gender. Specifically, female subjects may benefit more from the text plus verbal instructions than do male subjects. The length of the experiment was approximately 16 hours spread across 4 days.

**Results of Study:** Initial analysis indicates a gender by treatment interaction, with the female subjects benefiting more than the male subjects from the text combined with speech instructional treatment. In addition, the subjects within the speech treatment tended to agree that: (a) the speech made the task more interesting, (b) speech forced them to look at the displays, and (c) they learned more from the speech than if they had just read the screen. None of the subjects agreed with the statement that the speech was hard to understand.

**Selected Publications:** None.

**Keywords:** Modality, training, synthetic speech, spatial, gender, ATI

### **Phoenix: Testing the Effectiveness of Skill Acquisition (Joystick vs. Keyboard)**

**Scientists:** Goettl and Regian

**Instructors:** Linda Robertson-Schüle

**Total Subjects:** N = 28

**Purpose of Study:** The goal of this study was to examine the effectiveness of skill acquisition for the Phoenix desktop-flight simulator using either the joystick or keyboard only. Subjects trained on slalom and strike Phoenix tasks. The overall length of the study was approximately 4 hours, with military recruits used as subjects rather than civilians.

**Results of Study:** The effects that were obtained were largely practice effects. Although the joystick group tended to make more gates during training and the posttest, the results were generally not significant. However, there was a significant interaction between fidelity and practice for number of gates made on the hard courses. The keyboard group showed more improvement than did the joystick group. The cause may be the variability of practice effect. If this were the case then we would expect the joystick group to be better on a retention test or a transfer test; which was not measured, but rather a transfer test in the strike task. As in the slalom task, the performance trends favored the keyboard group (shorter average time, more kills, shorter distance, higher scores),

although the effects were not significant. It is possible that the study was not long enough and that the joystick group was still adjusting to the shifts from joy to key and back to joy and the effect will manifest itself later in training. Perhaps there was some compensatory rivalry among the subjects. It is possible those subjects in the joystick condition were not exposed to enough consistent mapping during training. The study will be rerun to delineate these factors.

**Selected Publications:** None.

**Keywords:** Fidelity, joystick, keyboard, spatial, training

### **Drive Simulator (Pilot)**

**Scientists:** Gugerty and Shute

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 70

**Purpose of Study:** This initial pilot study was to collect timing data, establish protocols, and to debug software. To develop a better understanding of "situational awareness" relative to the knowledge and cognitive processes involved in this complex construct. To answer quantitative questions regarding working memory, e.g., how many moving objects can people monitor and how accurately can they monitor them. Furthermore, to answer some qualitative questions, e.g., what type of strategies (based upon experience, i.e., novice/expert driver/pilot) does one implement when a crisis situation presents itself?

**Results of Study:** Timing data, experimental protocols, and software problems noted for correction.

**Selected Publications:** None

**Keywords:** Driving Simulator, Situational Awareness

### **Stat Lady II: Piloting the Descriptive Statistics Module of the Tutor**

**Scientists:** Shute

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 80

**Purpose of Study:** The purpose of the study was to gather pilot data on the instructional efficacy for the non-student model version of the Descriptive Statistics Tutor.

**Results of Study:** Pilot data gathered led to changes in the instruction and program abilities for future studies.

**Selected Publications:** None.

**Keywords:** Stat Lady Descriptive Statistics Tutor, ITS, ATI

### **System T<sup>2</sup>WO: Tutor vs. Traditional Classroom Instruction for Individuals vs. Dyads**

**Scientists:** Connolly, Kline, and Regian

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 109

**Purpose of Study:** Research conducted in the CoLab comparing traditional versus tutor-based learning using Stat Lady revealed no posttest differences. Significant differences were found when test items were categorized as declarative and procedural information. This finding suggests difference in underlying knowledge structure based upon the instructional method used. In addition, research findings have shown that collaborative learning is an effective, cost-efficient method of conducting training for skilled tasks. However, little is known about the effectiveness of collaborative learning for other types of tasks that are less skill specific. The purpose of this study was to evaluate the effectiveness of the tutor versus traditional classroom instruction for individuals versus dyads. The study consisted of approximately 20 hours distributed over five days. Post hoc analysis using Pathfinder/Multi-Dimensional Scaling techniques (MDS) will be used to evaluate possible differences in the underlying cognitive structure of the treatment groups as compared to that of the expert.

**Results of Study:** Initial results indicate that the tutor group performed significantly better than the traditional classroom group on the posttest. Specifically, the classroom group increased their performance by 1 standard deviation (8%) and the tutor group by 2 standard deviations (17%) on the posttest. Additionally, questionnaire results suggest that the tutor group preferred their instruction, thought that they had learned more, and were more open to taking the course again than were the classroom group.

**Selected Publications:**

Connolly-Gomez, C. (1994). US Air Force compares computer based training to traditional classroom training. Rappn' with Roscoe Newsletter, 1(1), p. 2.

**For more information contact:** Cathy Connolly at (941) 275-7720, CConnolly@galaxyscientific.com, or Kevin Kline at (210) 671-2667.

**Keywords:** Instructional method, collaborative, pedagogy, dyad, intelligent tutor

### **Team Composition and Type of Worked Examples in Algebra Word Problem Solving**

**Scientists:** Yadrick and Regian

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 102

**Purpose of Study:** The purpose of this study was to examine the simultaneous and/or interactive effects of difficulty of instructional approach and team composition of mathematics learning. Subjects worked individually or in teams, and were classified as high, medium, or low ability (relative to the subject population) and teamed with others based on their performance on the pretest and a section of the Armed Services Vocational Aptitude Battery (ASVAB). There were High-High, Low-Low, Medium-Medium, and High-Low ability teams. In addition, team composition was crossed with task complexity which was manipulated by presenting two different kinds of worked example sets. The 'difficult set' consisted of two relatively simple examples, leaving subjects to figure out how to adapt formulas and build equations for more complex problems. The 'easy set' consisted of a set of comprehensive examples, representing the full range of problem difficulty. The study consisted of approximately 16 hours over five days.

**Results of Study:** Awaiting further data.

**Selected Publications:** None.

**Keywords:** Problem solving, mathematics, ATI, example, collaborative

### **Phoenix: Investigating the Optimal Ratio of Practice to Observation for Learning**

**Scientists:** Goettl, Connolly, and Regian

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 97

**Purpose of Study:** Results in Observational Phoenix I were similar to those found in the Space Fortress dyadic study. Specifically, results indicated that participants showed superior performance over those in the control condition, reinforcing the theory that a complex task can be partially learned through observation. Moreover, comparisons of slalom and strike task performance revealed that participants acquired both strategic, as well as, psychomotor aspects of the task through observation. This study attempted to expand these findings and investigate the optimal ratios of practice to observation. Subjects either performed 100, 75, 50, 25, or 0 percent of the time or observed a partner for the other part of the time. All tests were taken individually. Based upon the results of this study, we attempted to determine if there was a basic underlying formula to guide practitioners in estimating how much observation time equates to "hands-on" time. The length of the task was approximately 16 hours across five days.

**Results of Study:** This study examined the performance to observation ratio in training a complex computer-based flight simulator task. Subjects were divided into five different groups. One group performed 100% of the time during training, the second group performed 75% of the time and observed 25% of the time, the third group performed 50% of the time and observed 50% of the time, the fourth group performed 25% of the time and observed 75%, and the fifth group observed 100% of the time. Based on previous observational learning literature, we predicted a linear relationship between performance and percent of time performing. Results however, suggest a non-linear relationship between performance and percent of time performing. Performance was slightly better when a combination of performance and observation was used rather than performance alone. These results indicate that observational learning plays an important, yet often neglected role in learning complex computer-based tasks and suggests that small group oriented computer-based training systems may be more instructionally and cost effective than individually oriented computer-based training systems.

#### **Selected Publications:**

Gomez, C.C., & Goettl, B. P. (1995). Investigating tradeoffs between practice and observation in automated instruction. *Proceedings of the 39th Annual Meeting of the Human Factors & Ergonomics Society* (1340-1344). Santa Monica, CA: The Human Factors & Ergonomics Society.

**Keywords:** Spatial, observation, cognitive strategies, modeling

## Investigating Three Theories of Symptom Elaboration

**Scientists:** Hale and Regian

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 60

**Purpose of Study:** Using the fictional industrial system called a glass brick maker (GBM) this experiment attempted to explore the role of symptom elaborations on the search component of symptom-fault learning. Two hypotheses were evaluated: (1) symptom elaboration improves search performance through topographic indexing, if original system learning is organized topographically, (2) symptom elaboration improves search performance through malfunction indexing if original system learning is organized functionally. A third hypothesis is that symptom elaboration improves search, independently of the nature of original learning.

**Results of Study:** Two measures were of interest in this experiment: (1) percent correct at each trial of rule learning and (2) trials to criterion. The first measure represents an absolute measure of accuracy and allows evaluation of the locus of the elaboration effect (hypothesis generation versus hypothesis recognition). The second measure allows an evaluation of the effect of symptom elaboration on learning rate. The percent correct results demonstrated a strong interaction between the type of information available during original learning and the nature of symptom elaboration during rule learning. Results showed that any type of symptom elaborations improved learning rates following original learning based on topography. Furthermore, there was no difference between topographic and malfunction symptom elaborations. Topographic elaborations significantly improved performance, topographic elaborations significantly degraded performance. These results suggest that providing learners with elaborative information, at the point of hypothesis generation, that allows them to construct causal chains from symptoms to the most plausible faults will greatly enhance learning by minimizing the number of errors committed during the early stages of rule construction.

### Selected Publications:

Hale, C.R. (November, 1994). *Explanations improve the search component of symptom-fault rule learning*. Annual meeting of the Psychonomic Society.

Hale, C.R. (October, 1994). *Training implications of differential elaboration in learning to troubleshoot devices*. Annual meeting of the Human Factors and Ergonomics Society.

**Keywords:** Troubleshooting, diagnosis, feedback, elaboration



### **Comparing Two Software Versions (Pilot Study)**

**Scientists:** Hale and Regian

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 11

**Purpose of Study:** Using Topo (version 7) software as a comparison, a newly developed version of the Topo software (version 8) was piloted to discern any differences between the software programs. Any programmatic type errors or faults will be changed and the newer version of the program will be used in the Topo 6 experiment.

**Results of Study:** Results revealed that the newly developed version eight of the Topo software did not contain any programmatic type errors and was more flexible, reduced the amount of time and effort to load the software and decreased the typographical and entry errors made by the test administrators.

**Selected Publications:** Awaiting follow-up study.

**Keywords:** Software, comparisons, Topo

### **Stat Lady II: Piloting the Data Organization and Plotting Module (DS-1) Criterion Tasks**

**Scientist:** Shute

**Instructors:** Connolly, Gawlick-Grendell, and Robertson-Schüle

**Total Subjects:** N = 79

**Purpose of Study:** The four forms of the Data Organization and Plotting Module (DS-1) Criterion Tests (i.e., Forms A, B, C, D) were pilot-tested for inter-item reliability. This pilot was used to determine whether items assessing the same Curricular Element (CE) were correlated with each other. Test item numbers are the same across the four forms (e.g., item 7 assessed CE-3) and, in the absence of instruction, subjects should perform similarly across matched / parallel items. Problems in forms A and B were designed to be isomorphic, in that they share similar formats. Questions in forms C and D were designed to be different in format than questions in A and B, yet isomorphic to one another. Subjects were assigned randomly to one of eight test orders that resulted from a fully counterbalanced design. The eight orders were: ACBD, ADCB, BDAC, BCDA, CBAD, CADB, DBCA, DABC.

**Results of Study:** Data analyses showed that the following combinations of tests are to be used in any upcoming experiments: A/D and B/C. Thus, half the subjects per study will receive forms A & D as their pretest and forms B & C as their posttest. The other half will receive forms B & C as their pretest and forms A & D as their posttest.

**Selected Publications:**

Shute, V.J. (1995). SMART: Student modeling approach for responsive tutoring. *User Modeling and User-Adapted Instruction (UMUAI)*. Special issue on "Student Modeling," (5) 1-44.

**Keywords:** Stat Lady Descriptive Statistics Tutor, criterion tests, interitem reliability

**Stat Lady II: Testing the Validity of Stat Lady as a Computer-Based Tutor**

**Scientist:** Shute

**Instructors:** Connolly, Gawlick-Grendell, and Robertson-Schüle

**Total Subjects:** N = 72

**Purpose of Study:** The purpose of the study was to determine how well subjects learned from this version of the tutor, and test the predictive validity of the computed p(CE) values. In addition, this pilot was used to repair any bugs in the program and acquire an estimate of learning times. Subjects were administered the non-student model version of the Descriptive Statistics Module 1 (DS-1) with remediation and the mastery criterion disabled (i.e., all subjects received only 2 problems to solve per problem set). Although the student model (SMART) was in place and updating itself (i.e., computing all p(CE)'s), it did not affect the course of the curriculum. Instruction time was fixed to a maximum of 3.5 hours. Subjects were also administered a demographic questionnaire and an affective survey to examine opinions about the tests, the tutor, and how well they believed they learned Descriptive Statistics.

**Results of Study:** Results showed about a 67% increase from Pretest to Posttest scores. Subjects appeared to learn very well from the tutor, and this becomes the score to beat in subsequent experiments involving the student model. Stepwise regression equations were computed predicting posttest from pretest, tutor data (i.e., p(CE)'s), and other demographic data (e.g., number of math courses, attitude about statistics, gender, education). In all instances, the tutor data was the strongest predictor of outcome, accounting for 34 to 63.4% of unique variance. Analyses of the affective survey revealed that subjects really enjoyed the instruction. Initially, most felt frustrated by the on-line pretest, but after exposure to the tutor, they reported that they mostly felt in control of

their learning, the tutor's feedback was helpful, and that they learned something. Based on the results of this pilot, the decision was made to allow more time on the tutor (i.e., > 3.5 hours) for the formal studies (i.e., Studies C and D). Numerous changes were also made to the program to clarify instruction and improve functionality.

**Selected Publications:**

Shute, V.J. (1995). SMART: Student modeling approach for responsive tutoring. *User Modeling and User-Adapted Instruction (UMUAI)*. Special issue on "Student Modeling," (5) 1-44.

**Keywords:** Stat Lady Descriptive Statistics Tutor, predictive validity, microadaptation, affect

**Investigating how Students Learn Troubleshooting Skills**

**Scientist:** Hale

**Instructors:** Connolly, Gawlick-Grendell and Robertson-Schüle

**Total Subjects:** N = 51

**Purpose of Study:** This study investigated issues surrounding how subjects learn to carry out a simple form of troubleshooting using a fictional industrial system called a glass brick maker (GBM) with the newest version of Topo (version 8). Subjects were presented five diagrams which characterize the system. The diagrams were presented individually and reflect a hierarchy comprised of top-level system diagrams followed by four diagrams representing subsystems (which decompose from the top level). Based on the treatment group, subjects received either explanatory information describing the functionality, describing the mechanism, or no explanation of each diagram component. In the learning phase, subjects studied and recalled simple diagrams. In the troubleshooting phase, subjects were required to generate fault hypotheses in response to symptoms. Retention tests were given 24, 48, or 96 hours following the learning phase.

**Results of Study:** Currently being analyzed.

**Selected Publications:** None.

**Keywords:** Software, comparisons, Topo

## **ATM 2 -- Task differences in Alternating Task Modules**

**Scientists:** Yadrick, Goettl, Connolly, Regian, and Shebilske

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 85

**Purpose of Study:** The purpose of this study was to see whether the similarity of the tasks influenced the magnitude of the ATM effect. This was the first of at least three projected experiments intended to examine the limits of Alternating Task Modules in massed vs. spaced training. Previous studies established that ATMs produce sizable spacing effects when there are three dissimilar tasks, one reasonable and important empirical question concerns how dissimilar tasks need to be in order to produce spacing effects. If for example, alternating Space Fortress with Phoenix produces spacing effects, will alternating Space Fortress with Space Fortress keyboard produce spacing effects? This question has practical implications. For example, in a military technical school it may be impractical to train by alternating two or three different subjects matters, since the purpose of the school is to train a single subject matter. As another example, in industry it may be difficult to train a given group of employees by alternating two or three completely different training topics. However, in either case it might be possible to train by alternating different aspects or areas of a topic or subject matter. In this study we will compare a group that alternates between Space Fortress and Phoenix with a group that alternates between Space Fortress and Space Fortress-Keyboard. About 15 subjects will be randomly assigned to each of 6 groups.

**Results of Study:** Currently being analyzed.

### **Selected Publications:**

Goettl, B. P., Yadrick, R.M., Connolly-Gomez, Regian, J. W., & Shebilske, W.L. Alternating task modules in isochronal training of complex tasks. *Human Factors*, 38(2), 330-346.

**Keywords:** Alphabet Arithmetic, Phoenix, Automaticity, Spatial, Quantitative

## **The Effects of Exercise on the Spacing Effect (A Control Group)**

**Scientists:** Goettl, Yadrick, and Connolly

**Instructors:** Connolly and Robertson-Schüle

**Total Subjects:** N = 76

**Purpose of Study:** The purpose of this study was to investigate the effects of exercise on the spacing effect using three different tasks: Space Fortress, Math tutor, and Phoenix. A previous study suggested that exercise reduces overall performance and size of the practice effect. In that study, exercise was subtracted from the total practice time. This study represented the comparison group for the exercise condition in the previous study. Participants repeated the massed and distributed schedules with exercise from the original massed versus distributed experiment. However, instead of exercising these subjects, they merely rested during the time allotted for exercise. Participants received 10 hours of training on Space Fortress, 16 hours on the Math Tutor, and 10 hours on Phoenix. There were 30 participants per group, two groups drawn from the temporary pool. About 30 participants were randomly assigned to each group with the constraint that the male:female ratio was constant across groups. The experiment took approximately 40 hours per week to complete. The most recent versions of Space Fortress, Math Tutor, and Phoenix were used.

**Results of Study:** Preliminary results indicated that the alternating task modules resulted in better performance on SF acquisition and retention. In addition, the exercise condition produced worse performance on the SF transfer task. Results for Algebra word problems and Phoenix are still being examined.

#### **Selected Publications:**

Goettl, B. P., Shebilske, W.L., Connolly-Gomez, C., & Robertson-Shule, L. (1996). The effects of exercise and alternating task modules in skill acquisition. Proceedings of the Human Factors and Ergonomics Society 40<sup>th</sup> Annual Meeting, 2, 1097-1101.

**Keywords:** Spatial, observation, cognitive strategies, modeling

### **Stat Lady II -- Descriptive Statistics Module (Computer-Based Tutor) Study D: Predictive Validity of SMART**

**Scientist:** Shute

**Instructors:** Connolly, Gawlick-Grendell, and Robertson-Schüle

**Total Subjects:** N = 104

**Purpose of Study:** The purpose of this study was to test the efficacy of the Stat Lady tutor and the predictive validity of SMART. In addition, comparisons between this study and Study C (microadaptive version of Stat Lady) were made to evaluate the remediation component (i.e., learning gains, time on tutor). Subjects were administered the non-student model version of the Descriptive Statistics Module 1

(DS-1) with remediation and the mastery criterion disabled (i.e., all subjects were limited to completing exactly 2 problems per problem set). Although, the student model (SMART) was in place and updating itself (i.e., computing all p(CE)'s), it did not affect the course of the curriculum. Subjects were also administered a demographic questionnaire and an Affective Survey (to examine opinions about the tests, the tutor, and how well they believed they learned Descriptive Statistics). In addition, cognitive and personality measures were administered and the Big Five Inventory was administered at the beginning and end of testing to measure test-retest reliability.

**Results of Study:** Pretest to posttest gains were found to be 2.2 standard deviations (Pre  $\bar{M}$  = 44%, SD = 14; Post  $\bar{M}$  = 75%, SD = 15). Stepwise regression equations were computed predicting posttest from pretest, tutor data (i.e., p(CE)'s), an aptitude factor score, education (years in school), and gender. The tutor data were found to be the first (and strongest) predictor of outcome, accounting for 54% of unique variance. Next to enter the equation was aptitude, accounting for an additional 11% of unique outcome variance. Analyses of the Affective Survey showed that initially, subjects were frustrated with the on-line pretest. After exposure to the tutor, they reported that they liked the tutor, enjoyed their learning experience, felt the tutor feedback was helpful, and felt in control of their learning. Preliminary comparisons with Study C data showed that learning efficiency was greater using the non-intelligent version of Stat Lady, but acquisition was significantly higher using the intelligent version.

#### **Selected Publications:**

Shute, V.J. (1995). SMART: Student modeling approach for responsive tutoring. *User Modeling and User-Adapted Instruction (UMUAI)*. Special issue on "Student Modeling," (5) 1-44.

Shute, V.J. & Gluck, K.A. (1995, April). *Artifacts of learning: The Stat Lady experience*. Paper presented at the annual meeting of American Educational Research Association, San Francisco, CA (April, 1995).

**Keywords:** Stat Lady Descriptive Statistics Tutor, microadaptation, macroadaptation, affect, Cognitive Abilities Measures (CAM), Big Five Personality Inventory

#### **Stat Lady II -- Descriptive Statistics Module (Computer-Based Tutor) Pilot C: Microadaptive Functioning of SMART (Debugging and Preliminary Results)**

**Scientist:** Shute

**Instructors:** Connolly, Gawlick-Grendell, and Robertson-Schüle

**Total Subjects:** N = 32

**Purpose of Study:** The purpose of this pilot was to test the functionality of SMART with mastery criterion, microadaptation, and remediation in place. In addition, this pilot was used to gain an estimate of tutor completion times with this version of the Stat Lady tutor and to test the stability of the program. Results from this pilot were used to determine an acceptable level for mastery. Subjects had to attain mastery of all Curricular Elements (CE's) before moving past any section of the curriculum. Otherwise, remediation was invoked. Instruction time was not fixed, to allow subjects all the time necessary to attain mastery of the curriculum (up to 2 full days). Subjects first took two on-line pretests, then received adaptive instruction; followed by two posttests. Subjects were also administered a demographic questionnaire and an Affective Survey (to examine opinions about the tests, the tutor, and how well they believed they learned Descriptive Statistics). In addition, cognitive and personality measures were assessed using a quantitative subset of the CAM 4.0 battery and the Big Five Personality Inventory.

**Results of Study:** The first group of subjects participating in this pilot was used to test various mastery criterion levels (i.e., 0.5, 0.6, 0.7, 0.8). It was found that 0.7 was the optimal level because the others made learning too easy or too hard. The decision rule used to determine *when* a student is remediated was also modified (based on whether/how often the student had been remediated on that CE previously). Based on the results of this pilot, the decision was made to allow more time on the tutor (i.e., up to 3 days rather than 2 days). Changes were also made to repair or improve the functionality of the program.

#### **Selected Publications:**

Shute, V.J. (1995). SMART: Student modeling approach for responsive tutoring. *User Modeling and User-Adapted Instruction (UMUAI)*. Special issue on "Student Modeling," (5) 1-44.

**Keywords:** Stat Lady Descriptive Statistics Tutor, mastery learning, microadaptation, affect, remediation, Cognitive Abilities Measures (CAM 4.0), Big Five Personality Inventory

#### **Assessing the Reliability of the Algebra Word Problem Solving (AWP) Pre/Post Tests**

**Scientists:** Yadrick , Regian, Connolly, and Robertson-Schüle

**Instructors:** Connolly, Gawlick-Grendell, and Robertson-Schüle

**Total Subjects:** N = 39

**Purpose of Study:** This study was conducted to assess the reliability of the (AWP) pre/post tests in response to reviewers' comments regarding a paper submitted for

publication. Paper-and-pencil tests were counterbalanced and administered in a 2-hour slot maximum, which was consistent with the original study.

**Results of Study:** Results revealed that the reliability of two forms of the math pretest and posttest were parallel.

**Selected Publications:**

Yadrick, R.M., Robertson-Schule, L., & Connolly-Gomez, C. (1996). Interface, instructional approach, and domain learning with a mathematics problem-solving environment. *Computers in Human Behavior*, 12(4), 527-548.

Yadrick, R.M., Regian, J.W., Connolly-Gomez, C., & Robertson-Schule, L. (1997). Dyadic vs. individual practice with exploratory and directive mathematics tutors. *Journal of Educational Computing Research*, 17(2), 165-186.

Yadrick, R.M., Regian, J.W., Connolly-Gomez, C., & Robertson-Schule, L. (1995). Individual and cooperative group learning with user-controlled and program-controlled mathematics tutors. (AL/HR-TP-1995-0031, ADA303591). Brooks AFB TX: Armstrong Laboratory, Human Resources Directorate.

Yadrick, R.M., Regian, J.W., Robertson-Schule, L. & Connolly Gomez, C. (1995). Interface, instructional approach, and domain learning with a mathematics problem-solving environment. (AL/HR-TP-1995-0032, ADA303561). Brooks AFB TX: Armstrong Laboratory, Human Resources Directorate.

**Keywords:** Algebra word problems, quantitative, reliability

**Testing the Number of Trials to Achieve Automaticity Utilizing  
the Phoenix and Alphabet Arithmetic Tasks**

**Scientists:** Yadrick, Shebilske, and Goettl

**Instructors:** Connolly, Gawlick-Grendell, and Robertson-Schüle

**Total Subjects:** N = 47

**Purpose of Study:** These pilots were conducted to prepare for the preliminary ATM 4 study. Subjects were given a series of Alphabet Arithmetic and Phoenix trials in order to estimate the total number of trials which could be completed within the time constraints and to work out procedures. Subjects were tested for a maximum of 4 hours on one day.



**Results of Study:** The results showed that 21 trials/sessions were sufficient for subjects to automatize the various problem sets within a 4-hour time slot.

**Selected Publications:** None. Awaiting follow-up study.

**Keywords:** Alphabet Arithmetic, Phoenix, automaticity, spatial, quantitative

### **Space Fortress: Assessing Gender Differences with Spatial Tasks**

**Scientists:** Regian and Shute

**Instructors:** Connolly, Gawlick-Grendell, and Robertson-Schüle

**Total Subjects:** N = 60

**Purpose of Study:** Initially, a pilot study was run to obtain timing data and to test various potential stress options. Subjects were trained on the Space Fortress task across two days to assess spatial gender differences during training.

**Results of Study:** Results found that sound (i.e., subjects listening to white noise while playing Space Fortress) did not impact upon subjects' performance. An uncertainty variable was identified to be used in subsequent analyses within Gender experiment #1. This was the pseudo-option version of Space Fortress (i.e., Ss were told that the Space Fortress version had changed, when actually it had not.). Also, there was no main effect found in the efficacy of group-discussions; in relation to improving female performance.

### **Selected Publications**

Regian, J.W. & Shute, V.J. (1994). Understanding and reducing gender differences in task performance. *Human Factors and Ergonomics Society Technical Training Group Newsletter, Spring* (1), 6-7.

**Keywords:** Gender differences, spatial, psychomotor

### **Stat Lady II -- Descriptive Statistics Module (Computer-Based Tutor) Study C: Microadaptive Functioning of SMART**

**Scientist:** Shute

**Instructors:** Connolly, Gawlick-Grendell, and Robertson-Schüle

**Total Subjects:** N = 107

**Purpose of Study:** The purpose of this study was to test the efficacy of SMART with mastery criterion, microadaptation, and remediation in place. Mastery was set at 0.70, and subjects had to attain mastery of relevant CE's (i.e., those below the pre-set mastery criterion) before moving on to the next section of the curriculum. Otherwise, remediation was invoked. Instruction time was not fixed, allowing subjects to attain mastery of the curriculum (up to 3 full days) at their own pace. Subjects first took the on-line pretest, then adaptive instruction, followed by the two posttests. Subjects also completed a demographic questionnaire and an Affective Survey (to examine opinions about the tests, the tutor, and how well they believed they learned Descriptive Statistics). In addition, cognitive and personality measures were assessed using a quantitative subset of the CAM 4.0 battery and the Big Five Personality Inventory.

**Results of Study:** Evaluating the remediation component involved comparing the relative degree of learning with this feature turned on (Study C) versus when it's not employed (Study D). Preliminary results of Study C showed significant gains in pretest to posttest learning (Pre  $\bar{M}$  = 50%; Post  $\bar{M}$  = 82%). Due to incoming differences on the pretest between Study C and Study D data, an ANOVA was computed on posttest using pretest as a covariate. Significant differences between groups were found, with higher performance levels being exhibited by Study C subjects. Thus, having the microadaptive features in place (i.e., remediation and mastery criterion) does help learners acquire greater degrees of domain-specific knowledge. In terms of learning efficiency, Study C subjects required significantly more time to complete the tutor (Study C  $\bar{M}$  = 7.6 hours; Study D  $\bar{M}$  = 4.4 hours). Thus, although the microadaptive instruction significantly increased acquisition, it also significantly added to learning time. Other analyses from these data are forthcoming.

### **Selected Publications:**

Shute, V.J. (1995). SMART: Student modeling approach for responsive tutoring. *User Modeling and User-Adapted Instruction (UMUAI)*. Special issue on "Student Modeling," (5) 1-44.

Shute, V.J. & Gluck, K.A. (1995, April). *Artifacts of learning: The Stat Lady experience*. Paper presented at the annual meeting of American Educational Research Association, San Francisco, CA (April, 1995).

**Keywords:** Stat Lady Descriptive Statistics Tutor, mastery learning, microadaptation, affect, remediation, Cognitive Abilities Measures (CAM 4.0), Big Five Personality Inventory

### **Early vs. Late Spacing of Training and its Effect on Automaticity**

**Scientists:** Yadrick, Shebilske, and Goettl

**Instructors:** Connolly, Gawlick-Grendell, and Robertson-Schüle

**Total Subjects:** N = 90

**Purpose of Study:** The purpose of the study was to determine effects of early or late spacing of training in an automaticity paradigm. One hypothesis was that massing training may be beneficial at particular points in the learning process. The purpose of this study was to follow acquisition of a new skill from its introduction until subjects have *automatized* the skill, using one of the four training schedules; either massed training throughout, an ATM regimen early in training followed by massed training, or an ATM regimen late in training following massed training. Two tasks were involved: Phoenix and an "Alphabet Arithmetic" task previously studied by Logan & Klapp (1991). In alphabet arithmetic, subjects verify correct or incorrect equations. For example:  $A + 2 = C$ , meaning that C is the second letter past A in the alphabet. Logan & Klapp's work using this paradigm supports their *retrieval theory of automatization*, wherein automatization is the process of *converting* from algorithmically calculating the answer for each problem to *retrieving* the answers directly from memory. Thus, this paradigm will allow differentiation of the effects of massing vs. ATM on different phases of performance. There was no test per se, because identifying an effect involves comparing reaction time and slope data for each session and group, to the corresponding data for each other session and group. We decided upon a filler task, Phoenix, which is a desktop-flight simulator. Subjects were tested once a week for a day, maximum 4-hour slot minimum.

**Results of Study:** Currently being analyzed.

**Selected Publications:** None.

**Keywords:** Alphabet Arithmetic, Phoenix, automaticity, spatial, quantitative

### **Space Fortress: The Effects of Gender, Testosterone Levels, Group Discussion, and Stress on Performance**

**Scientists:** Regian and Shute

**Instructors:** Gawlick-Grendell, Robertson-Schüle, Snooks, and Robbins

**Total Subjects:** N = 130

**Purpose of Study:** The purpose of this study was to investigate differences in training performance using Space Fortress (a complex spatial task) with regard to gender, testosterone levels, group discussion, and stress conditions. For

Gender Study experiment #1, subjects were trained for approximately 16 hours on the Space Fortress task (2-8 hour days). They completed nine training sessions on the Space Fortress task with five 15-minute discussion groups within training sessions. The Control group worked on a paper-and-pencil anagram task. Subjects that were tested had not played more than 20 hours per week of either home computer or video arcade games. In addition, they qualified on the Space Fortress Aiming Task (i.e., total score of 780 or more on at least one of three games). A signed consent form was obtained from the subjects for their sample of saliva. For Gender experiment #4, all subjects worked independently and completed four Space Fortress training sessions with the induced stress. Three stress conditions were used: (1) real options (i.e., more difficult SF version with subjects having to determine these changes on their own), (2) pseudo-options (i.e., subjects were the SF version had changed, when actually it had not), and (3) normal (i.e., subjects continued training with normal version of SF and were told nothing). Additionally, saliva samples were collected from each subject to evaluate testosterone levels.

**Results of Study:** Preliminary findings show that the relationship between testosterone levels and spatial performance were supported along with other findings in the literature concerning a significant quadratic trend.

**Selected Publications:** None.

**Keywords:** Gender differences, spatial, psychomotor

**Stat Lady II -- Descriptive Statistics Module (Computer-Based Tutor)**  
**Stat Lady E: Replication Study: Differential Practice Effects on Retention**

**Scientists:** Shute and Gluck

**Instructors:** Connolly, Gawlick-Grendell, Robertson-Schüle, Snooks, and Robbins

**Total Subjects:** N = 305

**Purpose of Study:** This replication study consisted of five practice conditions (between-subjects design) in DS-1: (a) AA (N = 80), (b) AE (N = 50), (c) EA (N = 50), (d) EE (N = 80), and (e) Learner's Choice (N = 80): (where A = Abbreviated, E = Extended, and LC = learners control number of problems they solve). Subjects in the Abbreviated condition solved one problem per problem set and in the Extended condition they solved four problems per set. The mixed groups switched over after problem set 3 ( $3/5 = .6$ ), as the original retention study switched mixed Ss after .61 of the curriculum was completed (14/23). Across five problem sets, the total number of problems to solve per condition were: AA (5), AE (11), EA (14), EE (20), LC (between 5-20). All

subjects were within the age range of 18-30 and had no prior Statistics courses in their educational backgrounds. Subjects first took an on-line pretest, received identical instruction from Stat Lady (between 1-3 days), then took an on-line posttest. All subjects will be recalled to participate in a follow-up study (6 months later), where they'll take a parallel test for assessment of retention (approximately 8 hours). Then the impact on learning (from self-selecting practice opportunities) will be assessed provided a projected *minimum* of 33% of all subjects to return.

**Results of Study:** Preliminary results showed that subjects assigned to the EE condition scored higher on their posttests in comparison to the other conditions. This was probably due to greater practice opportunities and problem-solving time. However, there were no differences with regard to extracurricular activities, (i.e., personal review of instructed material, accessing the on-line dictionary, and/or reviewing the hypertext reference guide defining symbols and formulas). Interestingly, results showed that subjects assigned to the LC condition not only learned the most from the Statistics tutor (reflected in obtaining the highest outcome scores), but they also acquired relevant knowledge and skills in significantly less time than the other conditions. This was apparent for both their instructional and problem-solving times relative to all other conditions.

#### **Selected Publications:**

Shute, V.J. (1995). SMART: Student modeling approach for responsive tutoring. *User Modeling and User-Adapted Instruction (UMUAI)*. Special issue on "Student Modeling," (5) 1-44.

Shute, V.J. & Gluck, K.A. (1995, April). *Artifacts of learning: The Stat Lady experience*. Paper presented at the annual meeting of American Educational Research Association, San Francisco, CA.

**Keywords:** Stat Lady Descriptive Statistics Tutor, microadaptation, macroadaptation, affect, Cognitive Abilities Measures (CAM), Big Five Personality Inventory

#### **Phoenix: Gender Differences Due to Spacing or Observational Effects?**

**Scientists:** Goettl and Connolly

**Instructors:** Gawlick-Grendell, Robertson-Schüle, Snooks, and Robbins

**Total Subjects:** N = 105

**Purpose of Study:** Our previous studies showed that subjects benefited from observing a Phoenix task that has significant cognitive processing demands. In that study females, not males, showed an observational learning effect. In a second

study, we found a direct relationship between the performance to observance ratio and performance, but that performance was best for a ratio of 75/25. In that study both males and females benefited from observational learning. The goal of this particular study was to further explore possible gender differences of observational learning and to disentangle spacing effects from observation effects. Subjects were placed into one of five training groups. All groups received training on the Phoenix strike task in one of the following schedules: (1) Spaced Control - alternate every two trials between Phoenix and filler task. The filler task used was Windows Solitaire. (2) Spaced Observation - alternate every two trials between perform and observe. (3) Massed Control - block of filler task followed by block of Phoenix. (4) Massed P-O - block of performing Phoenix, block of observing Phoenix. (5) Massed O-P - block of observing Phoenix, block of performing Phoenix. Results of the first two studies were promising. They showed support for Carroll and Bandura's model of observational learning. However, this study is needed to test specific control conditions to confirm our hypotheses. The design addressed the question of whether observation should be intermixed with opportunities to perform. In addition, it addressed the question of whether early observation is more important than late observation. Also, it was of interest to examine spatial and quantitative individual differences in observational learning, so CAM and VSAT batteries were obtained. The study took 2 days to complete.

**Results of Study:** Currently being analyzed.

**Selected Publications:** None.

**Keywords:** Gender differences, Observational learning, spacing, spatial, psychomotor

### **Assessing Individual Differences Using Individual Goals**

**Scientists:** Regian, Greenis, and Goettl

**Instructors:** Gawlick-Grendell and Robertson-Schüle

**Total Subjects:** N = 73

**Purpose of Study:** The purpose of the study was to investigate individual differences by assigning individualized goals to subjects. Results are expected to indicate ways in which goals may be incorporated into complex task training and Intelligent Tutoring Systems to maximize learning. This study serves as a predecessor to a forthcoming study, which will investigate the beneficial effects of process feedback and continuously updated, individualized goals on complex task performance. Research has shown that goal setting consistently leads to higher levels of performance on simple tasks. However, on complex tasks, setting goals does not result in increased performance and may result in

performance decrements. The goal of this study was to investigate one way in which goal setting may augment complex task performance, through the assignment of individualized goals. Subjects were placed into one of three training groups: (1) Control group, (i.e., their feedback read “do your best”), (2) Standard Goal, (i.e., subjects were given a difficult, specific performance goal, which was based on the average performance of individuals in the control group, and (3) Individualized Goal, (i.e., subjects were assigned a goal which was individually tailored to their initial performance and aptitudes. All subjects were between the ages 18 and 30, had an educational minimum of a high school or general equivalency diploma (GED), and had no prior job or experience with Air Traffic Control. Each subject performed 2 practice trials and nine performance trials (3 blocks of 3- 10 minute trials). Goals were updated before each new block of trials (every 30 minutes). Individualized goals are expected to facilitate performance by increasing self-efficacy and reducing negative affect and strategy switching.

**Results of Study:** Currently being analyzed.

**Selected Publications:** None.

**Keywords:** Spatial, Air Traffic Controller, varied goal feedback

### **Space Fortress: Examining Top-Down Processing**

**Scientists:** Shebilske and Goettl

**Instructors:** Gawlick-Grendell, Robertson-Schüle, Snooks, and Robbins

**Total Subjects:** N = 36

**Purpose of Study:** The purpose was to examine the role of top-down processing during intercession intervals in the distributed practice effect that has been observed during the acquisition of Space Fortress. We are specifically testing two hypotheses. First, we predicted that elaboration would be more effective later in skill acquisition when resource load has been reduced. Second, we predicted that the benefits of elaboration would be obtained primarily by subjects having a high elaborative cognitive style. To test these predictions we ran four different groups: (1) Massed Total Emphasis, (2) Massed Multiple Emphasis on Components, (3) Elaboration Total Emphasis, and (4) Elaboration Multiple Emphasis on Components. Each group was divided into high and low elaborators using a median-split technique. We predicted an interaction such that high-elaborators would benefit the most from the elaboration. This experiment provided an important signpost for our programmatic research on spacing effects and on observational learning in group protocols. The connection to observational learning was through the

hypothesis that it also facilitated learning through top-down processing. All subjects had no prior exposure to other Space Fortress experiments.

**Results of Study:** Currently being analyzed.

**Selected Publications:** None.

**Keywords:** Elaborations and Individual Differences

### **Phoenix Observational Learning III**

**Scientist(s):** Barry P. Goettl & Cathy Connolly-Gomez

**Instructors/Proctor(s):** {Instructor/Proctor Name(s)}

**Total Subjects:** N = 98

**Purpose of Study:** The goal of this study was to explore whether massed or spaced observation was effective and when was the best time to apply observation in training. Five groups were compared using a desktop flight simulator task. Massed-perform-observe (MPO) group performed a block of massed trials, while massed-observe-perform (MOP) group observed. MPO and MOP groups switched their functions after one hour. The participants in spaced-observe (SO) group alternated between practice and observation every two trials. These three groups were compared to massed- and spaced-control conditions. Massed observation was expected to be more effective because it provides participants with more opportunities to develop a mental representation of the task prior to the actual task performance.

**Results of Study:** In contrast to the spaced observation group, participants in the massed observation group performed the task better than those in the control conditions. Also, participants in the massed performance group performed the task better than participants in the massed-control condition. These results suggest that observational learning is beneficial for both, the observer and performer.

**Selected Publications:**

Goettl, B. P., & Connolly-Gomez, C. (1997). The effects of observational learning and practice schedule on the acquisition of a complex skill. Proceedings of the Human Factors and Ergonomics Society 41<sup>st</sup> Annual Meeting (1180 – 1184). Santa Monica, CA: The Human Factors & Ergonomics Society.



**Keywords:** Phoenix, observational learning, massed and spaced practice schedule, complex skill acquisition

### **Air Traffic Control 1 – Hold One**

**Scientist(s):** Barry P. Goettl, F. L. Lee and J.R. Anderson

**Instructors/Proctor(s):** Michael Anthony

**Total Subjects:** N = 111

**Purpose of Study:** The goal of this experiment was to examine the relationship between aptitude, strategy, and performance in the Kanfer-Ackerman Air Traffic Control (ATC) task as the use of strategy changes with the change in ATC task environment. Manipulated variables were the taxiing time of the planes on runways and the penalty for using lower holds strategy versus no such penalty. In contrast to long taxiing time, short taxiing time and no penalty condition were expected to increase the utility of Hold 1 strategy.

**Results of Study:** As predicted the use of strategy changed as the implicit and explicit changes in the ATC task environment altered. A strong positive relationship was found between aptitude and performance. Also, a strong relationship was found between keying rate and performance, while no penalty condition showed a strong negative relationship between Hold 1 strategy and performance. Most of the strong relationship between cognitive ability and performance was through the relationship between ability and Hold 1 strategy.

### **Selected Publications:**

Lee, F. L., Anderson, J. R., & Goettl, B. P. (1997). The relationship between aptitude and strategy use in performing dynamic, real-time tasks. Available in <http://act-r.psy.cmu.edu/publications/pubinfo>

**Keywords:** Air Traffic Control task, adaptive strategy, aptitude and performance

### **Space Fortress Dyad-VEC**

**Scientist(s):** Barry P. Goettl, Michael Anthony, and Wayne L. Shebilske

**Instructors/Proctor(s):** {Instructor/Proctor Name(s)}

**Total Subjects:** N = 168

**Purpose of Study:** The goal of this study was to create more efficient and effective protocol by combining dyadic protocol (individual vs. dyad) and emphasis change manipulation (component vs. total score) on the Space Fortress computer task. Also examined were the effects of gender and interaction anxiety. Moreover, the relationship between performance on the explicit and implicit task components and consequent verbalizable knowledge (IE questionnaire) was examined.

**Results of Study:** The results suggested that participants in dyads improved more on ship control than individuals; women benefited more from emphasis change than men and performance of individuals high in interaction anxiety was inhibited when they were in the dyad-VEC condition. Moreover, performance on explicit task components was positively related to the consciousness ratings of explicit components while performance on implicit tasks and consciousness ratings of implicit components showed near-zero correlations.

**Selected Publications:**

Goettl, B. P., Anthony, M., Derek, K., Snooks, S., & Shebilske, W. L. (1998). Emphasis, gender and interaction anxiety in individual and dyadic training protocols. Proceedings of the Human Factors and Ergonomics Society 42<sup>nd</sup> Annual Meeting (1412-1416). Santa Monica, CA: The Human Factors & Ergonomics Society.

Goettl, B. P., Anthony, M., Derek, K., & Shebilske, W. L. (1998). Consciousness ratings correlate with explicit but not implicit components of a complex task. Paper accepted for presentation at the 39<sup>th</sup> Annual Meeting of the Psychonomics Society.

**Keywords:** Space Fortress, variable emphasis on components, explicit/implicit dissociation

**Air Traffic Control 2 – Individual Differences**

**Scientist(s):** Barry P. Goettl and Christian Schunn

**Instructors/Proctor(s):** Michael Anthony

**Total Subjects:** N = 57

**Purpose of Study:** The goal of this experiment was to introduce another approach to individual differences in cognitive skill which states that people differ in how well they can adaptively change their strategies as a result of changing features of the task environment. The task used in this experiment was Kanfer-Ackerman Air Traffic Control task. Also, CAM-4,

a cognitive assessment battery was used to determine what cognitive abilities are associated with adaptiveness.

**Results of Study:** The results indicated that most participants adapted their choice of strategy in response to the alterations in the task environment. One of the cognitive sub-skills, inductive reasoning skill, was associated with adaptiveness in the general context of the ATC task and variable manipulations used in this experiment.

#### **Selected Publications:**

Reder, L. M., & Schunn, C. D. (1998). Bringing together the psychometric and strategy worlds: Predicting adaptivity in a dynamic task. In D. Gopher and A. Koriath (Eds.), Attention and Performance XVII: Cognitive regulation of performance: Interaction of theory and application. Cambridge, MA: MIT Press.

**Keywords:** ATC task, individual differences, adaptive strategies

#### **Aircraft Identification: Orientation Constancy in the Picture Plane I**

**Scientist(s):** Ashworth and Robbins

**Proctor(s):** Robbins

**Total Subjects:** N = 67

**Purpose of Study:** To quantify the normalization function associated with identification of military aircraft that are rotated in the picture plane. To quantify the normalization differences as a function of overall visual similarity of the stimulus set.

**Results of Study:** Line drawings of current military aircraft of both American and Foreign origin served as stimuli. Similarity rating of pair-wise presented aircraft and card sorting tasks were used to cluster aircraft into highly similar (homogeneous) or highly distinctive (heterogeneous) stimulus sets. Both stimulus sets were presented to subjects in an RSVP identification paradigm wherein the subjects first learned to associate individual planes with individual response keys, and then were instructed to identify planes that were systematically rotated within the picture plane. Resulting orientation normalization functions resembled classical "mental rotation" curves, with homogeneous stimuli producing curves of significantly high slope than heterogeneous stimuli. Results are interpreted to support the hypothesis that highly homogeneous stimuli force the object recognition system to use quantitative features for identification (which are more

orientation specific) while highly heterogeneous stimuli force the use of more qualitative features (which are less orientation specific).

**Selected Publications:**

Ashworth, A.R.S. III, & Robbins, R. D. (1996). Visual homogeneity affects orientation sensitivity. The 37th Annual Meeting of the Psychonomic Society, Chicago, IL.

**Keywords:** Aircraft Identification, Orientation Constancy, Mental Rotation

**Aircraft Identification: Orientation Constancy in the Picture Plane II**

**Scientist(s):** Ashworth and Robbins

**Proctor(s):** Robbins and Snooks

**Total Subjects:** N = 81

**Purpose of Study:** Replicates and extends the results of Orientation Constancy in the Picture Plane I to three-dimensional realistically rendered military aircraft.

**Results of Study:** Three-dimensional realistically rendered stimuli of current military aircraft of both American and Foreign origin served as stimuli. Similarity rating of pair-wise presented aircraft were used to cluster aircraft into highly similar (homogeneous) or highly distinctive (heterogeneous) stimulus sets. Both stimulus sets were presented to subjects in an RSVP identification paradigm wherein the subjects first learned to associate individual planes with individual response keys, and then were instructed to identify planes that were systematically rotated within the picture plane. Resulting orientation normalization functions resembled classical "mental rotation" curves, with homogeneous stimuli producing curves of significantly high slope than heterogeneous stimuli. Results are interpreted to support the hypothesis that highly homogeneous stimuli force the object recognition system to use quantitative features for identification (which are more orientation specific) while highly heterogeneous stimuli force the use of more qualitative features (which are less orientation specific).

**Selected Publications:**

Ashworth, A.R.S. III, & Robbins, R. D. (1996). Maintaining object constancy for recognition of aircraft. The 3rd Annual Recognition Materials Conference. Langley AFB VA: 480th Intelligence Group, 36 IS/Recognition Materials Branch.

**Keywords:** Aircraft Identification, Orientation Constancy, Mental Rotation

### **Space Fortress – MEC Plus Rules**

**Scientist(s):** Barry P. Goettl, Shirley F. Snooks, and Wayne L. Shebilske

**Instructors/Proctor(s):** {Instructor/Proctor Name(s)}

**Total Subjects:** N = 233

**Purpose of Study:** The goal of this study was to examine the relationship between multiple emphasis on components (MEC) and verbal elaboration. The participants were trained on the Space Fortress task. The assumption was that verbal elaboration would focus participants on the explicit processes while MEC would focus participants on the implicit processes. Four training conditions were developed: MEC and TOT (total score emphasis) with and without verbal elaboration.

**Results of Study:** The results showed that MEC protocol was more beneficial for the low-aptitude learners while TOT protocol was more beneficial for the high-aptitude learners. The results on the component scores showed that combination of MEC and verbal elaboration increased the scores on speed (explicit component) but decreased the scores on control (implicit) component.

#### **Selected Publications:**

Goettl, B. P., Snooks, S. F., & Shebilske, W. L. (1997). Emphasis change and verbal elaboration in skill acquisition: A tale of two components. Proceedings of the Human Factors and Ergonomics Society 41<sup>st</sup> Annual Meeting (1195-1199). Santa Monica, CA: The Human Factors & Ergonomics Society.

**Keywords:** Space Fortress, multiple emphasis on components, verbal elaboration

### **Space Fortress Observational Learning I**

**Scientist(s):** Barry P. Goettl and Julie Stiles-Shipley

**Instructors/Proctor(s):** {Instructor/Proctor Name(s)}

**Total Subjects:** N = 264

**Purpose of Study:** The goal of these studies was to look at the effects of observation (no observation, novice model and experienced model observation), model skill level (novice and experienced) and practice schedule (massed vs. spaced) on the acquisition of a complex computer-based task (Space

Fortress). The dependent measure was the average total point scores for test sessions.

**Results of Study:** The results suggested that there was a slight advantage for observing experienced models at the beginning of training although later in training there was no difference between experienced and novice model observation. Both, novice and experienced model observation proved to be more effective than no-observation condition. Moreover, males only seemed to have benefited more from observation of either novice or experienced models. Also, males showed a greater benefit of spaced training schedule than females.

#### **Selected Publications:**

Goettl, B. P., Stiles-Shipley, J., Derek, K., Jimenez, L., & Hardcastle, K. (1998). Observing expert and novice models in acquisition of complex skills. Proceedings of the 16<sup>th</sup> Applied Behavioral Sciences Symposium (121-125). US Air Force Academy, Colorado Springs, CO.

Stiles-Shipley, J. A., Goettl, B. P., Connolly, C., Jimenez, L. (1998). The influence of expert and novice models on acquisition of a complex skill. Proceedings of the Human Factors and Ergonomics Society 42<sup>nd</sup> Annual Meeting (1641). Santa Monica, CA: The Human Factors & Ergonomics Society.

**Keywords:** observational learning, practice schedules, human performance

### **Temporal Reasoning Study (pilot)**

**Scientist(s):** Lisa Torreano

**Instructors/Proctor(s):** Michael Anthony, Shirley Snooks

**Total Subjects:** N = 36

**Purpose of Study:** This pilot is an initial investigation of how people make inferences that depend upon temporal information, specifically the order of events occurring over time. After reading a four-sentence description about everyday events, participants are posed a question concerning the order of the occurrence of two of the events described in the scenario. The purpose of this pilot is to (a) measure the reading times for the premises and questions for each scenario, (b) record the accuracy of the inferences made, and (c) assess the types of incorrect inferences that are made.

**Results of Study:** No conclusions have been drawn from this data-set due to (a) intermittent failure of the programs to run properly, and (b) the difficulty that some of the participants had in doing the task; further analysis is required.

**Selected Publications:** N/A (pilot study)

**Keywords:** temporal reasoning

### **Validating Alternative Pedagogical Approaches to Teaching Basic Algebra Skills** (*Math2*)

**Scientist(s):** Ross Willis

**Instructors/Proctor(s):** Michael Anthony, Kristie Hardcastle, Leticia Jiminez

**Total Subjects:** N = 40

**Purpose of Study:** This study was a pilot study designed to test two novel pedagogical approaches to teaching students how to solve basic algebraic equations and word problems. The first graphical pedagogy is an analogy that likens solving an algebraic equation to maintaining balance on a seesaw. Then solving algebraic equations, one must be careful to perform an operation to both sides of the equation in order to maintain equality, or balance. A seesaw presented beneath an equation tilts when an operation has been performed to one side of the equation and not to the opposite side of the equation. The second graphical pedagogy is a “highlighter” method that successively highlights keywords in a word problem and dynamically builds an algebraic equation representing the situation described in the word problem.

**Results of Study:** Learners who used the seesaw pedagogy did not have higher posttest scores (controlling for pretest score) than learners who merely practiced solving algebraic equations, in the absence of instruction. However, learners who used the highlighter pedagogy in the word problems lesson achieved significantly higher posttest scores (controlling for pretest score) than participants assigned to the practice condition.

**Selected Publications:** N/A (pilot study)

**Keywords:** intelligent tutoring systems, algebra instruction

**A Comparison of Four Methods of Remediating Students' Errors  
in a Computer-based Instructional System**  
(*Math3*)

**Scientist(s):** Ross Willis

**Instructors/Proctor(s):** Michael Anthony, Kristie Hardcastle, Leticia Jiminez, John Billimek, Eduardo, Henry

**Total Subjects:** N = 169

**Purpose of Study:** Computer-based instruction may address a student's remedial needs along two dimensions: re-teaching format and remedial elements. In terms of re-teaching format, the computer-based instructional system can re-teach the deficient knowledge and skills using the initial instructional screens in the form a review of the original material. Alternatively, the system can present a set of new instructional screens based on an alternative pedagogical strategy (as described in Math2). In terms of remedial elements, the instructional system may present a complete set of instructional screens designed to address all elements of the curriculum, as in a complete review of the initial instructional materials. Alternatively, the system may diagnose the student's error and select a set of instructional screens specifically designed to address the student's apparent misconception. This experiment examined the four remediation methods and, additionally assessed individual differences with regard to the four remediation methods. The principal research question was: Which form(s) of remediation branching is (are) best for which learner(s)?

**Results of Study:** Results from Experiment 3 did not support the notion of aptitude-treatment interactions using the manipulations described previously. Overall, learners with high working-memory proficiency achieved higher posttest scores and required less time to complete the lessons than learners with low working-memory proficiency.

**Selected Publications:** None.

**Keywords:** intelligent tutoring systems, remediation, cognitive diagnosis, aptitude-treatment interactions



## **GK Comprehension**

**Scientist(s):** Lisa Torreano

**Instructors/Proctor(s):** Kristie Hardcastle, Henry Avila

**Total Subjects:** N = 72

**Purpose of Study:** This study is an initial step towards determining whether there are general cognitive abilities (e.g., that are identifiable w/a subset of the CAM tests) that correlate with patterns of processing metaphorical uses of verbs.

**Results of Study:**

**Selected Publications:** None.

**Keywords:** figurative language, metaphor processing

## **Building a Bug Library**

**Scientist(s):** Ross Willis

**Instructors/Proctor(s):** Ross Willis

**Total Subjects:** N = 28

**Purpose of Study:** When an instructional system utilizes cognitive diagnosis to determine a student's specific misconception, a bug library must be constructed. This was accomplished by having 28 participants complete a paper-and-pencil algebra quiz consisting of 34 problems. Participants were instructed to explicitly show all work used to compute answers.

**Results of Study:** I analyzed the written work of each participant who obtained incorrect responses for problems. Doing so, I was able to recognize several incorrect solution paths that were followed by one or more participants. These data were used in the final version of the instructional system.

**Selected Publications:**

Willis, R. E., & Taraban, R. (submitted). A problem-solving approach to analyzing students' algebra errors. Submitted to the Journal of Developmental Education.

**Keywords:** intelligent tutoring systems, cognitive diagnosis, bug library